

المعهد العربي للتخطيط بالكويت

مقدمة في

علم الإحصاء التطبيقي

تأليف
د. علي ابو القاسم محمد



مقدمة في

علـــم الاحصـــاء التطـــبيقي

تأليــف

د. علي أبو القاســـم محمـــد

بسم الله الرحمن الرحيم

توطئــة

يهدف هذا المؤلف الى تعريف الدارسين والقارئين بالتحليل الإحصائي وأدواته وبمختلف المقاييس والمؤشرات الإحصائية التي يمكن اشتقاقها من المجتمعات أو الظواهر موضوع التحليل .

ولقد أعد هذا المؤلف للدورة التمهيدية لدبلوم تخطيط التنمية بالمعهد العربي للتخطيط على اعتبار أن السادة المتدربين من خلفيات علمية متباينة ، فمنهم من سبق له دراسة الإحصاء في كليات التجارة أو الإقتصاد ومنهم من لم يسبق له التعرف على الطرق والأدوات الإحصائية لهذا فقد روعي في إعداده التبسيط والتطبيق ، ومن ثم فإننا افترضنا عدم وجود أي قدر من المعرفة لدى الدارس لنبني على أساسه طرق اشتقاق مقاييس وأدوات التحليل الإحصائي المطلوبة لأغراض هذه الدورة.

ولقد توخينا في طرقنا لمختلف الأدوات والمقاييس الترابط بين كل من مقاييس النزعة المركزية والتشتت وكذلك الأرقام القياسية . كما حاولنا تبسيط الإثبات الجبري تمكيناً للسادة المتدربين من الإستفادة القصوى ، لهذا جعلنا الإثبات الرياضي في حدود ضيقة واسترشدنا بما أمكن من الأمثلة والتطبيقات بقدر كاف تتحقق معه الغاية من التعلم والتدريب ، وأفردنا في نهاية كل فصل عدداً من التطبيقات العملية روعي في اختيارها الصلة بالمتغيرات الإقتصادية .

يود المؤلف أن يتقدم بجزيل الشكر للأستاذ الفاضل عبدالله محمد علي مدير المعهد العربي للتخطيط لتشجيعه على طبع هذا الكتاب كما يشكر السادة حسن الحاج وعبدالله على محمد على الجهد الذي بذلاه ليصير هذا الكتاب حقيقة واقعة بعد أن كان مجرد فكرة.

والله الموفـــق .

المؤلـــف

الكويت 1984

د. علي أبو القاسم محمـــد

المحتويـــات

مقدمـــة

لقد عرفت الإحصاءات من قديم الأزمنة حيث كانت تستخدم لأغراض حربية وضريبية وفلكية ، وازدادت أهميتها في القرن الثامن عشر وخاصة بعد نشوب الثورة الصناعية حينما أيقن رجال الأعمال من ضرورتها من أجل اتخاذ قرارات سليمة . إلا أن الاحصاء كعلم لم يظهر إلا في نهاية القرن الثامن عشر وكان أول من أرسى قواعده العالم كواتيله Quetlet (1796 - 1874) .

فالإحصاء - علم - له طرقه المختلفة وقوانينه المتعددة وأسسه الثابتة ونظرياته العلمية المرنة المتطورة ، وتطبيقاته الواسعة الإنتشار في مجال حياتنا العملية كما أن له علاقاته المتشعبة والمتبادلة بالعلوم الأخرى حيث يؤثر فيها ويتأثر بها . والإحصاء بمفهومه الحديث يخدم الباحثين في جميع الميادين العلمية ومتخذي القرارات في المجالات العملية . وعلى سبيل المثال فإن الباحث في مجال الإقتصاد يستطيع أن يختبر نظرياته عن سلوك المستهلك أو علاقة المستخدم- المنتج عن طريق استخدام الطرق الإحصائية . كما أن الباحث في مجال الطب يستخدم نفس هذا الأسلوب لقياس كفاءة دواء جديد أو لإيجاد علاقة بين التدخين ومرض معين . . . كما يستخدمها أيضاً الباحث في المجال الزراعي لمعرفة آثار الأسمدة المختلفة على محصول معين مثلاً . . . ويمكن القول عموماً أنه لا يوجد ميدانٌ من ميادين البحث العلمي إلا وطرقه علم الإحصاء ولعب دوراً كبيراً في تطوره . أما بالنسبة لمتخذ القرارات سواء كانت قرارات إدارية أو حربية فإنه لن يستطيع أن يستغنى عن الأساليب الإحصائية في دراسته للقرارات البديلة قبل اتخاذ قراره .

وسوف نتناول في هذا المقرر كل من أنواع الأساليب الإحصائية ، والمراحل

الأساسية في البحث الإحصائي والأخطاء الإحصائية .

فمن هذا المنطلق يمكن تقسيم الأساليب الإحصائية الى ثلاثة أقسام رئيسية :

1 ـ الإحصاء الوصفي Descriptive Statistics وهو يختص بوصف خصائص البيانات المستخدمة في البحث الإحصائي . فإذا كانت لدينا بعض البيانات الخاصة بظاهرة معينة ، فعلم الإحصاء الوصفي يبين لنا كيف يتم توزيع هذه البيانات وما إذا كانت تتمركز حول قيمة معينة أم أنها متباينة وإذا كانت هناك علاقة بين الظاهرة وظاهرة أخرى وما قوة هذه العلاقة .

2 ـ الإستدلال الإحصائي: Statististical Inference وهو يختص باستخلاص نتائج عامة من بعض المشاهدات . ويتم ذلك عن طريق أسلوب المعاينة الإحصائية Statistical Sampling أو أسلوب المعاينة كما يسمّيه البعض .

وتجدر الإشارة هنا الى تعريف كل من المجتمع والعينة ، يقصد بالمجتمع Population جميع المفردات موضع الدراسة والتي نريد معرفة حقائق عنها سواء كانت منها المفردات في شكل انسان أو حيوان أو جماد . فعلى سبيل المثال قد يكون لدينا مجتمع من سكان مدينة معينة أو مجتمع من الخيل أو مجتمع من درجات امتحان الطلبة في مادة معينة . . . ويقصد بالعينة Sample مجموعه من مفردات المجتمع مثال : عينة من سكان مدينة معينة أو عينة من درجات امتحان الطلبة في مادة معينة . . .

هذا ويختص أسلوب المعاينة الإحصائية بدراسة وتحليل مجموعة صغيرة من المفردات ـ أي عينة منها ـ حتى يتم الوصول الى نتائج يمكن تعميمها على مجتمع هذه المفردات بأسره . وفي مثل هذه الدراسات عادة تمثل العينة المستخدمة المجتمع تمثيلاً جيداً وأي « معلومة » تستنتج من العينة ينظر اليها على أنها تقدير للمعلومة الفعلية المناظرة لها ، أي المعلومة التي كان سيحصل عليها إذا ما تم تحليل ودراسة المجتمع بأسره .

هذا وتمكن الطرق الإحصائية الباحث من تحديد ما الذي يتوقعه من خطأ نتيجة لاستخدام الاستدلال الإحصائي .

3 ـ التنبؤ Forecasting ويقصد به استخدام المشاهدات الماضية للإستدلال بها لما سيحدث للظاهرة موضع البحث في فترة زمنية مقبلة . فإذا فرضنا أن لدينا علاقة خطية بين متغير س ومتغير آخر ص ، ولتكن ص هي المبيعات من سلعة معينة ، و س الزمن بالسنوات ، ولنفرض أننا نريد التنبؤ بمبيعات هذه السلعة في فترة زمنية مقبلة ، أن التنبؤ هنا يقوم على استخدام العلاقة بين المتغيرين للإستدلال على قيمة المتغير ص ـ أي كمية المبيعات ـ في فترة زمنية مقبلة استناداً الى استمرار العلاقة في المستقبل على ما كانت عليه .

الفصــل الأول

تجميع وتصنيف وعرض البيانات

GROUPING AND CLASSIFICATION OF DATA

الفصـــل الأول

تجميع وتصنيف وعرض البيانات

GROUPING AND CLASSIFICATION OF DATA

مقدمـــة :

تجمع البيانـات الإحصـــائية سواء عن طريق الحصر الشامـــل Complete Enumeration كما في حالة التعدادات Census كالتعداد السكاني ،أو عن طريق التجربة Experiment أو عن طريق العينات Samples .

وفي الغالب أن حجم هذه البيانات كبيـر وفي صورة غير منتظمة وبالتالي لا يمكن استخلاص المعـلومات والحقائق من هذه البيانات قبل تـنظيمها وتلخيصها في صورة جدولية أو بيانية أو تصويرية ، ومن أمثلة الرسوم البيانية الخط البياني العادي والخط البياني النصف لوغاريثمـي والخـط البيانـي اللوغاريثمي والأعمـدة البيانية Barcharts أمـا الرسـوم التصـويرية Pictographs فهـي رسـوم مبسطـة وتصلـح للصحف والمجلات وتناسب القارىء العادي .

الصفة المتغيرة : Variables

الصفة المتغيرة هي الصفة القابلة للتغير من فرد إلى آخر في المجتمع ويلاحظ أن معظم الصفات البيولوجية تعتبر صفات متغيرة . ويطلـق على الصفـة التـي تتغـير عشوائياً اسم المتغير العشوائي Random Variable القيمة التي تعطى لكل فرد من صفة معينة تعرف باسم متغير Variate ومن الصفات المتغيرة اللون والوزن والطول والحجم ، عدد الأزهار ، عدد الحشرات . . . الخ .

والصفات المتغيرة إما أن تكون صفات كمية Quantitative مثل الحجم ، الوزن ، الدخل ، عدد البتلات في الأزهار. أما الصفات المتغيرة الوصفية Qualitative فهي مثل اللون أو الطعم أو الجنس .

وتقسم الصفات المتغيرة من الوجهة الإحصائية الى صفات مستمرة Continuous Variables وصفات غير مستمرة أو متقطعةDiscrete Variables .

والصفة المستمرة هي التي يمكن أن تأخذ أي قيمة عددية في مدى معين . مثل الطول ، الوزن ، فمثلاً إذا كان طول نبات معين يتراوح ما بين 30 - 300سم نعني بذلك أنه يمكن نظرياً أن يتواجد نبات بأي طول في هذا المدى ولو أننا محددين بدقة وحدات القياس .

أما الصفة المتغيرة المتقطعة فهي تلك الصفة التي تتغير بوحدات كاملة مثال ذلك عدد البتلات في الأزهار فهي أما 1 ، 2 ، 3 ، . . . الخ . ولكن لا توجد زهرة ذات $2\frac{1}{2}$ بتله أي أنه لا توجد لها أرقام وسطية ـ كذلك من أمثلة الصفات المتقطعة حجم الأسرة ، عدد الحشرات في قطعة أرض،وقد يطلق عليها أحياناً الصفات العددية للبيانات (Enumeration Data) ويرمز عادة للصفة المتغيرة بالرمز س(X) أو ص (Y) . . . الخ .

مثـال :

إذا كانت الصفة المتغيرة تحت الداسة هي أوزان ستة من الرياضيين لذا فان X_i تمثل وزن الرياضي رقم (i) والجدول التالي يوضح أوزان الرياضيين الست .

رقم الرياضي	1	2	3	4	5	6
الرمـز	X_1	X_2	X_3	X_4	X_5	X_6
وزن الرياضي (رطل)	153	186	210	214	175	180

فقيمة X_1 ترمز الى وزن الرياضي الأول وهو 153 رطل وهكذا.

وكثيراً ما نستعمل الأحرف اليونانية في الإحصاء فمثلاً يرمز الى مجموع القيم بالرمز (Sigma Σ) ويقابلها مجـ ـ فإذا أردنا التعبير عن مجموع الأوزان للرياضيين الست فإن :

$$\sum_{i=1}^{6} X_i = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 1118$$

مصادر جمع البيانات :

إذا اعتمد الباحث في حصوله على البيانات المطلوبة للدراسة على بيانات قد تم تجهيزها ونشرها من جهة أخرى أطلق على المصدر في هذه الحالة مصدر غير مباشر ، أما إذا تم للباحث إعداد وتجهيز البيانات بصفته الشخصية أو من خلال الجهاز الفني القائم بالدراسة فإنه يطلق على المصدر في هذه الحالة مصدراً مباشر .

أولاً : المصدر الغير مباشر في الحصول على البيانات :

حيث نحصل على البيانات من مصادرها غير الأولية ، جاهزة ، مبوبة ومصنفة ودون أن يبذل الباحث أدنى مجهود في تكوينها ، ومن أمثلة ذلك النشرات والدوريات والمجلات والكتب العلمية والسجلات والمحفوظات والمذكرات وكل ما تنشره الجهات والهيئات ومراكز البحوث العلمية من تقارير وبحوث ودراسات إحصائية سابقة .

ومن أمثلة هذا النوع من المصادر أيضاً تلك النشرات والدوريات التي يقوم بنشرها الجهاز المركزي للإحصاء أو التقارير التي تنشرها الوزارات عن متابعة تنفيذ الخطة الإقتصادية للبلاد أو ما ينشر كل ثلاثة شهور عن النشاط المصرفي وما تنشره أجهزة الأمم المتحدة وغيرها .

ويلجأ الباحث الى هذا المصدر في جمع البيانات إذا ما وجد صعوبة في الحصول على البيانات بطريقة مباشرة من مصادرها الأولية أو رغبة في استكمال بقية

البيانات من نتائج بحوث سابقة تتعلق بموضوع الدراسة ، أو قد يرى الباحث أن هذا الأسلوب سيحقق له دقة مطلوبة أعلى مما لو اعتمد على أسلوب آخر .

ثانياً : المصدر المباشر في الحصول على البيانـات :

وهنا يعتمد الدارس عند الحصول على البيانات اللازمة للدراسة على المصادر الأولية لهذه البيانات ويقوم بتكوينها وتجهيزهابطريقة مباشرة ، دون الإعتماد على ما نشر قبل ذلك من بيانات أو تم تجهيزه لدى أي جهة أخرى.

ويلجأ الباحث الى هذا المصدر إذا كان من الصعب الحصول على بيانـات معدة جاهزة عن الدراسة أو أن طبيعة الدراسة تفرض إتباع هذا الأسلوب كما لو كان يريد استطلاع رأي مجموعه من المستهلكين عن رأيهم في نوعية جديدة من سلع الإستهلاك ، ولا شك في أن درجة الدقة والثقة في بيانات هذا المصدر إذا جمعت بطريقة دقيقة وصحيحة تكون مرتفعة مما يساعد على الوصول لنتائج موثوق فيها .

وعند مقارنـة المصدر المباشر بالمصدر الغير مباشر نجد أن المصدر الغير مباشر يفضل بسهولته وسرعة الحصول على البيانات من خلاله وتوفيره للوقت والجهد والتكاليف . بينما يعاب عليه صعوبة تحديد درجة الدقة والثقة في تلك المعلومات وعدم الحصول على البيانات بالطريقة التفصيلية اللازمة للدراسة ، الى جانب خطورة الإعتماد على أكثر من مصدر غير مباشر واحد لاختلاف أساس التكوين على نحو ما أسلفنا .

أما المصدر المباشر فإنه يساعد في الحصول على بيانات معروف درجة الثقة والدقة فيها وبالتالي يمكن تحديد هذه الدرجة عند تحليل البيانات كمياً وهي في الغالب ما تكون مرتفعة ويكون لها أساس تكوين واحد وبالتالي تصلح لعملية المقارنة بين متغيرات الظاهرة الواحدة الى جانب أن هذه البيانات مرنة يمكن تطويعها وتشكيلها بالطريقة المطلوبة لغرض البحث والدراسة .

إلا أنه من مشاكل الإعتماد على المصدر المباشر في تجهيز وتكوين المعلومات عن الظواهر المختلفة ، الحاجة الى الوقت الكافي لذلك . وربما لا يتمشى ذلك مع بعض الأبحاث التي يحتاج الدارس فيها الى تحديد نتائجها بسرعة ، كما أن هذا الأسلوب يلزمه جهاز فني كبير للحصول على المعلومات بطريقة مباشرة من حقل أو ميدان ، ناهيك عن الصعوبات التي تعترض الباحث أثناء جمع البيانات وما يلزم الدراسة من اعتمادات مالية خارجة عن قدرة الباحث المادية في كثير من الأحيان مما قد تبعده عن القيام بهذه الدراسة .

تعريفـــات :

أ ـ المجتمع : Population

يتكون المجتمع من كل القيم التي يمكن أن يأخذها فرد أو أنها تمثل مجموعة من الأفراد Individuals والتي تشترك في صفة واحدة أو أكثر ، مثال ذلك مجتمع الأبقار البلدية في السودان فهي تمثل جميع الأبقار البلدية في السودان ويمكن دراسة هذا المجتمع بالنسبة لصفة واحدة أو أكثر ، فيمكن دراسة الوزن ، كمية إدرار اللبن ، الحجم ، اللون . فيمكن تكوين مجتمع يمثل أوزان الابقار وهي عبارة عن الأوزان المختلفة لجميع الأبقار في السودان . وقد يمثل المجتمع مجموعة القياسات الممكنة لفرد واحد مثال ذلك إذا أردنا حساب وزن حيوان معين فعند أخذ وزن عدد من هذه الحيوانات فإن الأوزان المتتالية لن تعطي نفس الوزن بل أنها قد تختلف من حيوان الى آخر ، ولو أن معظم هذه الأوزان سوف تلتف حول قيمة واحدة هي متوسط المجتمع ، والمجتمع أما أن يكون من النوع المحدود Finite أي أنه يمكن حصر جميع أفراده مثال ذلك مجتمع من أطفال مدينة ما . وقد يكون المجتمع غير محدود Infinite ومن أمثلة ذلك مجتمع من البكتريا في كمية من اللبن أو عدد النقط التي يتكون منها خط ما أو عدد القياسات الممكنة لمساحة ما ، وعموماً تعامل المجتمعات على أنها غير محدودة إذا صعب حصر أفرادها .

تعرف المقاييس الإحصائية الخاصة بالمجتمع والمميزة له باسم **ثوابت** (معالم) المجتمع Parameters مثل متوسط المجتمع والإنحراف المعياري . ومن النادر حساب مثل هذه الثوابت « المعالم » من المجتمع وذلك لصعوبة حصر جميع أفراد المجتمع ولكنها عادة تستنتج من العينات المأخوذة من المجتمع .

ب ـ العينة : Sample

العينة هي جزء من المجتمع أخذت لتمثل المجتمع أو أنها قد تكون من جميع أفراد المجتمع كما في حالة المجتمع الصغير المحدود . وقد تكون العينة الطريقة العملية لدراسة المجتمع وذلك في حالة المجتمعات الكبيرة وذلك لصعوبة حصر جميع أفراد المجتمع أو لزيادة تكاليف الفحص ، وقد تدرس العينة لاستحالة فحص المجتمع حيث أن عملية الفحص تتسبب في إتلاف الأفراد المفحوصة ، مثال ذلك عند تقدير نسبة الحموضة في منتجات مصنع معلبات فإن عملية الفحص تتطلب إتلاف أفراد العينة تحت الفحص .

وتدرس العينة بهدف عمل استنتاجات إحصائية عن المجتمع ولذلك يجب أن تكون العينة ممثلة للمجتمع فمثلاً يجب مراعاة أن تمثل العينة جميع أفراد المجتمع وألا تكون العينة متحيزة Biased لجزء من المجتمع .

ويجب أن تتوفر في العينة الممثلة Representative Sample الشروط التالية :

1 ـ يجب أن تكون أفراد العينة ممثلة للمجتمع تحت الدراسة .

2 ـ يجب ألا تكون الأفراد المختارة تمثل قسم معين من أقسام المجتمع بل يجب أن تمثل جميع أفراد المجتمع .

وتعرف المقاييس المحسوبة من العينة باسم إحصائيات أو مقاييس العينة Statistics وهي تقدير Estimate لما يقابلها من ثوابت (معالم) المجتمع التي تمثله .

الرموز العلمية: Notations

عند كتابة أي رقم وخاصة إذا كان متضمناً عدداً كبيراً من الأصفار قبل أو بعد العلامة العشرية ، فإنه من المفيد استخدام الرمز العلمي للأساس .

مثـال 1 -

$10^2 = 10 \times 10, \quad 10^5 = 10 \times 10 \times 10 \times 10 \times 10, \quad 10^8 = 100000000$

مثـال 2 -

$10^0 = 1, \quad 10^{-1} = 0.1, \quad 10^{-2} = 0.01, \quad 10^{-5} = 0.00001$

مثـال 3 -

$864000000 = 8.64 \times 10^8, 0.00003416 = 3.416 \times 10^{-5}$

لاحظ أن ضرب رقم بـ 10^8 مثلاً يؤدي الى تحريك العلامة العشرية 8 أماكن الى اليمين . كما أن ضرب رقم بـ 10^{-6} يؤدي الى تحريك العلامة العشرية 6 أماكن الى اليسـار .

من المعتاد أن تستخدم الأقواس أو النقط للتعبير عن ضرب رقمين أو أكثر ، مثلاً

$5 \times 3 = 15$ تساوي $5.3 = 15$

$10 \times 10 \times 10 = (10)(10)(10) = 10.10.10 = 1000$

إذا استخدمت الحروف للدلالة على أرقام فإنه من المعتاد حذف الأقواس أو النقط. على سبيل المثال:

$ab = (a)(b) = a.b = a \times b$

وتعد الرموز مفيدة في العمليات الحسابية وتستخدم قاعدة

$(10)^p)(10^q) = 10^{p+q}, \frac{10^p}{10^q} = 10^{p-q}$

حيث p و q أي رقمين

في الرقم 10^p ، p تسمى الاس و10 الاساس

مثـال 1 ـ

$$(10^3)\ (10^2) = 1000 \times 100 = 100000 = 10^5 \text{ (i.e. } 10^{3+2}\text{)},$$

$$\frac{10^6}{10^4} = \frac{1000\ 000}{10000} = 100 = 10^2 \text{ (i.e. } 10^{6-4}\text{)}.$$

مثـال 2 ـ

$$(4\ 000\ 000)\ (0.000\ 000\ 0002) = (4 \times 10^6)\ (2 \times 10^{-10})$$

$$(4)\ (2)\ (10^6)\ (10^{-10}) = 8 \times 10^{6-10} = 8 \times 10^{-4}$$

مثـال 3 ـ

$$\frac{(0.006)\ (80\ 000)}{0.04} = \frac{(6 \times 10^{-3})\ (8 \times 10^4)}{4 \times 10^{-2}} = \frac{48 \times 10^1}{4 \times 10^{-2}}$$

$$\left(\frac{48}{4}\right) \times 10^{1-(-2)} = 12 \times 10^3 = 12\ 000$$

رمز الدليل : Subscript

الرمز X_j (يقرأ X «دليل j ») يمثل أي من القيم $X_1, X_2, X_3, \ldots, X_n$ والتي يأخذها المتغير X وعددها n . الحرف j في X_j الذي يمكن أن يكون أي رقم 1, 2, 3,..., n يسمى الدليل أو الرقم الجانبي الأسفل. ومن الواضح أن أي حرف آخر غير j مثل i, k, p, q, s يمكن أيضاً استخدامه .

رقم التجميـــع :

الرمز $\sum_{j=1}^{n} X_j$ يستخدم للدلالة على مجموع كل الـ X_j s ابتداء من $j = 1$ الى $j = n$ بالتعريف .

وإذا لم يكن هناك أي غموض محتمل فإننا نعبر عن هذا المجموع بشكل أبسط بالرمز $\sum_j X_j$ or $\sum X_j$ ، الرمز Σ هو حرف التاج اليوناني سيجما ونعني به هنا المجموع .

مثال 1

$$\sum_{j=1}^{n} X_j Y_j = Y_1 X_1 + Y_2 X_2 + Y_3 X_3 + \ldots . Y_n X_n$$

مثال 2 ـ

$$\sum_{j=1}^{n} aX = aX_1 + aX_2 + \ldots + aX_n$$

$$= a(X_1 + X_2 \ldots + X_n) = a\sum_{i=1}^{n} X_j$$

حيث a ثابت ـ وبشكل أبسط

$$\Sigma aX = a\Sigma X$$

مثال 3 ـ إذا كانت a, b, c ثوابت ، أثبت أن

$$\sum_{j=1}^{n} (aX_j + bY_j - CZ_j) = a\sum_{j=1}^{n} X_j + b\sum_{j=1}^{n} Y_j - C\sum_{j=1}^{n} Z_j$$

الحـــل :

$$\sum_{j=1}^{n} (aX_j + by_j - CZ_j)$$

$$= (aX_1 + bY_1 - CZ_1) + (aX_2 + bY_2 - CZ_2) + \ldots + (aX_n + bY_n - CZ_n)$$

$$= (aX_1 + aX_2 + \ldots + aX_n) + (bY_1 + bY_2 + \ldots + bY_n) - (CZ_1 + CZ_2 + \ldots + CZ_n)$$

$$= a(X_1 + X_2 + \ldots + X_n) + b(Y_1 + Y_2 + \ldots Y_n) - C(Z_1 + Z_2 + \ldots + Z_n)$$

$$= a\sum_{j=1}^{n} X_j + b\sum_{j=1}^{n} Y_j - C\sum_{j=1}^{n} Z_j$$

أو باختصار

$$\sum (aX + bY - CZ) = a\sum X + b\sum Y - C\sum Z$$

التوزيعات التكرارية : Frequency Distribution

البيانات الخام Raw Data : هي بيانات جمعت ولكنها غير منظمة عددياً . مثال ذلك مجموعة أوزان 100 طالب استخرجت من سجلات جامعة حسب الترتيب الأبجدي لأسمائهم .

المفردات المنظومة Arrays : المنظومة هي ترتيب للبيانات الرقمية الخام ترتيباً تصاعدياً أو تنازلياً حسب قيمتها . الفرق بين الرقم الأكبر والرقم الأصغر يسمى مدى **Range** البيانات . على سبيل المثال ، إذا كان أكبر الطلبة وزناً في المائة طالب هو 74 كلغم وأقلهم وزناً هو 60 كلغم . فإن المدى هو 74 - 60 = 14 كلغم .

الجداول التكرارية : Frequency Tables

عند تلخيص أعداد كبيرة من البيانات الخام فإنه من المفيد توزيعها على فئات أو طوائف وتحديد العدد الذي ينتمي لكل فئة ويسمى هذا العدد بتكرار الفئة والجدول المنظم على صورة فئات يقابل كل فئة تكرارها يسمى بالتوزيع التكراري أو الجدول التكراري .

مثـــال :

الاوزان (كيلوجرامات)	عدد الطلبـــة
62 ـ 60	5
65 ـ 63	18
68 ـ 66	12
71 ـ 69	27
74 ـ 72	8
المجمــوع	100

يمثل الجدول أعلاه توزيع تكراري لأوزان (مقربة الى أقرب كلغـم) 100 طالب من طلبة احدى الجامعات .

الفئة أو الطائفة الأولى على سبيل المثال تشتمل على الأوزان من 60 كغم الى 62 كلغم . ويعبر عنها بالرمز 60 ـ 62 . وبما أن عدد الطلبة الذين ينتمون الى هذه الفئة هم 5 طلبة فإن التكرار المقابل لهذه الفئة هو 5 .

تسمى البيانات المنظمة والملخصة كما في التوزيع التكراري أعلاه بالبيانـات المجمعة وعلى الرغم من أن عملية التجميع تؤدي بشـكل عام الى ضياع كثـير من تفصيلات البيانات الأصلية فإن الفائدة الهامة منها هي الصورة العامة التي يمـكن الحصول عليها والعلاقات الأساسية التي تظهر أكثر وضوحاً .

فترة الفئات وحدود الفئـات : Class Interval & Limits

الرمز الذي يعبر عن الفئة مثل 60 ـ 62 في الجدول أعلاه يسمى بفترة الفئة .

الرقمان 60 و62 يسميان حدود الفئة . الرقم الأصغر 60 يسمى الحد الأدنى للفئة ، والرقم الأكبر 62 يسمى الحد الأعلى للفئة والمصطلح فئة أو فترة يستخدمان في أغلب الأحيان للدلالة على نفس المعنى.

والفئة التي ليس لها حد أعلى أو حد أدنى تسمى فئة مفتوحة . على سبيل المثال إذا أخذنا مجموعة أعمار لأشخاص فإن الفترة (65 سنة فأكثر) هي فترة مفتوحة.

الحدود الحقيقية للفئات Class Boundaries

إذا كانت الأوزان قد سجلت الى أقرب كلجم فإن الفئة 60 ـ 62 تتضمن من الناحية النظرية كل القياسات من 59.5 كلجم الى 62.5 كلجم اذن الرقم الأصغر 59.5 هو الحد الأدنى الحقيقي للفئة والرقم 62.5 هو الحد الأعلى الحقيقي للفئة .

ومن الناحية العملية فإن الحدود الحقيقية للفئة يمكن الحصول عليها بجمع الحد الأعلى للفئة والحد الأدنى للفئة التالية لها والقسمة على 2 .

طول الفئة : Class Length

طول الفئة هو الفرق بين الحد الأدنى الحقيقي والحد الأعلى الحقيقي للفئة . إذا كانت جميع الفئات في التوزيع التكراري لها نفس الطول فإن الطول المشترك يرمز له بالرمز C . وفي هذه الحالة فإن C هو الفرق بين الحدين الأدنيين لفئتين متتاليتين . أو الحدين الأعليين لفئتين متتاليتين .

مركز الفئة : Class Mid — point

مركز الفئة هو منتصف الفئة ونحصل عليه بجمع الحد الأدنى والحد الأعلى

للفئة ونقسم المجموع على 2 . فمركز الفئة 60 - 62 هو $\frac{62+60}{2}$ أي 61 ويسمى مركز الفئة أيضاً بمنتصف الفئة . وبهدف المزيد من التحليل الرياضي فإنه يفترض أن جميع القراءات الموجودة داخل فترة فئة تأخذ قيماً تتطابق مع مركز الفئة . بهذا فإن جميع الأوزان داخل الفئة 60 ـ 62 تعتبر كما لو أنها 61 كلجم .

قواعد عامة لتكوين التوزيعات التكرارية :

1 ـ حدّد أكبر قيمة وأقل قيمة في البيانات الخام ومنها أوجد المدى (الفرق بين أكبر وأصغر رقم) .

2 ـ قسم المدى الى عدد مناسب من الفئات المتساوية الطول . إذا لم يكن ذلك ممكناً استخدم فئات ذات أطوال مختلفة أو فئات مفتوحة . ويأخذ عدد الفئات عادة بين 5 و12 حسب البيانات . وتختار الفئات أيضاً بحيث يتفق مركز الفئة مع المشاهدات الفعلية . وهذا يؤدي الى التقليل من أخطاء التجميع عند إجراء مزيد من المعالجة الرياضية . وعلى أية حال فإن الحدود الحقيقية للفئات يجب أن لا تتفق مع بيانات مشاهدة فعلاً .

3 ـ حدّد المشاهدات التي تقع في كل فئة . أي حدد تكرار كل فئة وأحسن طريقة لأداء ذلك هو استخدام كشف الحزم Tally أو النقط.

تمريـــن

(1) درجات 80 طالباً في مادة الرياضيات في جامعة ما مسجله بالجدول التالي:

93	76	88	62	90	68	82	75	84	68
75	85	59	71	93	60	73	88	79	73
72	63	78	95	62	74	87	75	65	61
60	68	74	69	77	94	75	82	78	66
71	83	79	60	95	75	61	89	78	96
75	71	65	76	85	78	97	67	62	79
74	53	76	62	78	88	57	73	80	65
77	85	75	76	63	72	81	73	67	86

بالرجوع للجدول أعلاه حدّد:

(أ) أكبر درجة (ب) أقل درجة (جـ) المدى (د) درجات أعلى خمسة طلبة من حيث الترتيب (هـ) درجات أقل خمسة طلبة من حيث الترتيب (و) درجة الطالب الذي ترتيبه العاشر من أعلى (ز) ما هو عدد الطلبة الذين حصلوا على الدرجة 75 فأكثر (ح) ما هو عدد الطلبة الذين حصلوا على درجات أقل من 85 . (ط) ما هي النسبة المئوية للطلبة الحاصلين على درجات أعلى من 65 ولكن ليست أعلى من 85 (ي) ما هي الدرجات التي لم تظهر مطلقاً .

(2) في الجدول التالي سجلت أطوال 40 طالباً من طلبة إحدى المدارس الى أقرب سم . كون توزيعاً تكرارياً .

157	149	125	144	132	150	164	138
144	152	148	136	147	140	158	146
165	154	119	163	176	138	126	168
135	140	153	135	147	142	173	146
128	145	156	150	142	135	145	161

***

العرض البياني : Graphic Representation

المدرجات التكرارية والمضلعات التكرارية : هما طريقتان في الرسم البياني للتعبير عن التوزيعات التكرارية .

1 ـ المدرج التكراري أو مدرج التكرارات يتكون من مجموعة من المستطيلات لها : ـ

أ ـ قاعدة على المحور الأفقي (محور X) مركزها عند مركز الفئة وطول القاعدة يساوي طول الفئة .

ب ـ مساحة متناسبة مع تكرارات الفئات

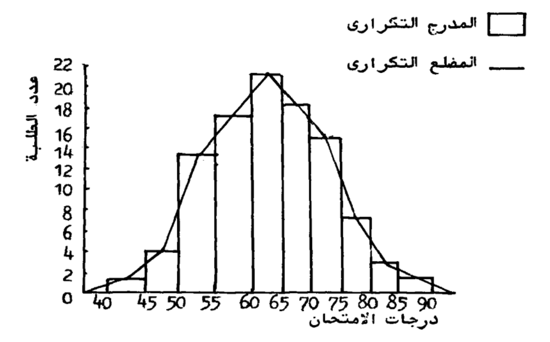
وإذا كانت الفئات كلها لها نفس الطول فإنه من المعتاد أن تأخذ الارتفاعات مساوية لتكرارات الفئات ، أما إذا كانت الفئات غير متساوية الطول فإن هذه الأطوال يجب أن تعدل .

2 ـ المضلع التكراري : هو خط بياني لتكرار الفئة المقابل لمركز الفئة . ويمكن رسمه بإيصال نقط تنصيف رؤوس المستطيلات المكونة للمدرج التكراري .

مثـــال:

الجدول أدناه يوضح التوزيع التكراري لدرجات امتحان 100 طالـب في مادة الإحصـاء.

درجات الامتحان	عدد الطلبة
أقل من 40	صفـــر
40 -	1
45 -	4
50 -	13
55 -	17
60 -	21
65 -	18
70 -	15
75 -	7
80 -	3
85 وأقل من 90	1
المجموع	100



فيما تقدم بينّا كيفية رسم المدرج التكراري في حالة الفئات المتساوية ، فإذا كانت الفئات غير متساوية فإنه من الخطأ تمثيل التكرارات كما هي على المحور الرأسي ، بل يجب تعديلها قبل رسم المدرج التكراري والسبب في ذلك يرجع الى أن مساحة كل مستطيل تمثل التكرارات ، فعندما تكون الفئات متساوية تكون مساحة المستطيل متناسبة مع التكرار المناظر ، ولكن في حالة الفئات الغير متساوية يختلف الوضع . وبصفة عامة يمكن القول بأن مساحة المستطيل تمثل التكرار .

∵ مساحة المستطيل = القاعدة × الارتفاع

= طول الفئة × الارتفاع

$$\therefore \text{الارتفاع} = \frac{\text{مساحة المستطيل}}{\text{طول الفئة}}$$

ويعبر عن ارتفاع المستطيل بالتكرار المعدل أي أن

$$\text{التكرار المعدل} = \frac{\text{التكرار} \times \text{طول الفئة الثابت}}{\text{طول الفئة المراد تعديلها}}$$

مثـــال:

فيما يلي توزيع أعمار سكان قرية معينة حسب السن ، والمطلوب رسم المدرج التكراري لهذا التوزيع .

فئات السن	أقل من سنة	1 -	5 -	10 -	20 -	30 -	40 -	50 -	60 فأكثر
عدد السكان	100	140	160	140	200	180	150	70	50

الحـــل:

الملاحظ أن جدول التوزيع التكراري السابق مفتوح من أعلى ومن أسفل وفي الواقع يمكننا قفله من أسفل فنكتب الفئة الأولى كالآتي (0 -) ، بينما لا يمكن قفله من أعلى ، كما أن من الملاحظ أن الفئات غير متساوية فطول الفئة الأولى 0.95، سنة بينما طول الفئةالثانية 3.55 سنة وطول الفئة الثالثة 5 سنوات ، وطول الفئة الرابعة 10 سنوات ، لذلك قبل رسم المدرج التكراري يجب إيجاد التكرار المعدل حيث أن

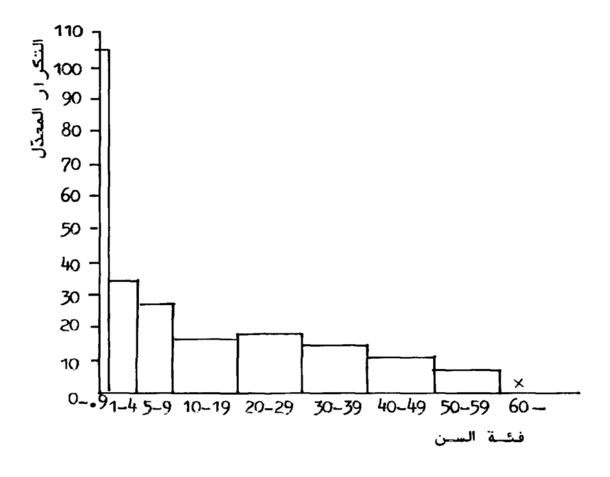
$$\text{التكرار المعدل} = \frac{\text{التكرار} \times \text{طول الفئة الثابت}}{\text{طول الفئة المراد تعديلها}}$$

الجدول أدناه يوضح التكرار المعدل لتوزيع السكان في القرية .

فئة السن	عدد السكان	طول الفئة (سنة)	التكرار المعدل
صفر - 0.95	100	0.95	105
0.95 - 4.5	140	3.55	39.4
4.5 - 9.5	160	5	32
9.5 - 19.5	190	10	19
19.5 - 29.5	200	10	20
29.5 - 39.5	180	10	18
39.5 - 49.5	150	10	15
49.5 59.5	70	10	7
59.5 -	50	—	—

الشكل أدناه يوضح المدرج التكراري المعدل لسكان القرية

(على اعتبار سنة واحدة هو الطول الثابت)

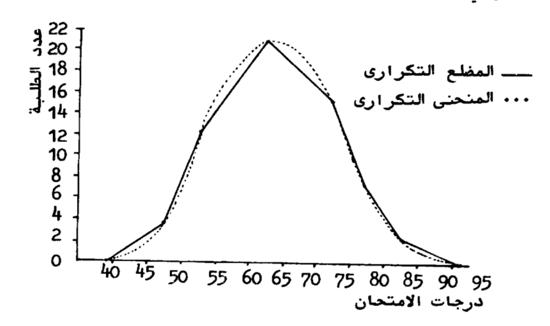


نلاحظ في هذا الشكل أن الفئة الأخيرة غير ممثلة لأنها فئة مفتوحة وهنا يكفي وضع علامة × فوق هذه الفئة للتدليل على أنها فئة مفتوحة .

المنحنى التكراري : Frequency Curve

بعد رسم المدرج التكراري يمكن تمهيد المنحنى التكراري باليد بحيث يتوسط المرور بين النقط الممثلة لرؤوس المضلع التكراري حتى يكون شبه انسيابي وليس خطوط منكسرة كما في حالة المضلع . هذا ويمكن القول أنه كلما زاد عدد المفردات في العينة كلما أمكن تصغير أطوال الفئات وكلما أقترب المضلع التكراري من المنحنى التكراري .

فإذا أخذنا الجدول السابق في المثال الذي يوضح التوزيع التكراري لدرجات امتحان 100 طالب في مادة الاحصاء فمنه يمكن رسم المضلع والمنحنى التكراري .



المنحنى التكراري المتجمع : Cumulative Frequency Curve

1 ـ التكرار المتجمع الصاعد :

إذا كان لدينا توزيعا تكراريا وأردنا معرفة عدد المفردات التي تكون قيمتها أقل من قيمة معينة نوجد ما يسمى « بالتكرار المتجمع الصاعد » .

ولتوضيح ذلك إذا نظرنا للجدول السابق الذي يمثل جدول التوزيع التكراري

لدرجات امتحان 100 طالب في مادة الاحصاء ، وإذا أردنا معرفة كم طالب حصل على أقل من 40 درجة فسيكون الجواب : ليس هناك طلاباً حصلوا على أقل من 40 درجة . وإذا أردنا معرفة كم طالب حصل على أقل من 45 درجة سيكون الجواب : طالب واحد ، وإذا أردنا معرفة كم طالب حصل على أقل من 50 درجة سيكون الجواب (1 + 4 = 5) ، وبالمثل فإن عدد الطلبة الذين تقل عدد درجاتهم عن 55 هو (1 + 4 + 13 = 18) . ولعمل جدول التوزيع التكراري المتجمع الصاعد نتبع الخطوات الآتية .

١ ـ نضيف الى الفئات فئة قبل الفئة الأولى وسيكون التكرار المناظر لها صفر . فمثلاً في الجدول السابق يمكن إضافة فئة قبل الفئة الأولى وهي : أقل من 40 وسيكون التكرار المناظر لها صفر.

2 ـ نرسم عمودين : الأول يبين « أقل من الحد الأعلى للفئة » والثاني خاص بـ « التكرار المتجمع الصاعد » على النحو المبين في الجدول أدناه.

3 ـ نحسب التكرار المتجمع الصاعد ، فبالنسبة لأقل من 40 يكون التكرار المتجمع الصاعد = صفر . ثم بعد ذلك نضيف التكرار المتجمع الصاعد الى تكرار الفئة التالية ، فيكون التكرار المتجمع الصاعد 1 ، ثم 5 ثم 18 . . . وهكذا الى أن نصل الى أقل من الحد الأعلى للفئة الأخيرة وهنا يجب أن يكون التكرار المتجمع الصاعد مساو لمجموع التكرارات .

ولتوضيح ذلك انظر الجدول التالي :

جدول التوزيع التكراري والتكرار المتجمع الصاعد لدرجات ١٠٠ طالب في مادة الإحصاء .

التكرار المتجمع الصاعد النسبي	التكرار المتجمع الصاعد	أقل من الحد الأعلى للفئة	التكرار	الفئة
0	0	اقل من 40	0	أقل من 40
0,01	1	أقل من 45	1	40 -
0,05	5	أقل من 50	4	45-
0.18	18	أقل من 55	13	50-
0,35	35	أقل من 60	17	55-
0.56	56	أقل من 65	21	60-
0.74	74	أقل من 70	18	65-
0,89	89	أقل من 75	15	70-
0.96	96	أقل من 80	7	75-
0.99	99	أقل من 85	3	80-
1.00	100	أقل من 90	1	85 وأقل من 90
			100	المجموع

ويمكن إيجاد التكرار المتجمع الصاعد النسبي بقسمه كل تكرار متجمع صاعد على جموع التكرارات كما في العامود الأخير من الجدول أعلاه .

٢ ـ التكرار المتجمع النازل :

إذا أردنا معرفة عدد المفردات التي تكون قيمتها مساوية وأكبر من قيمة معينة نوجد التكرار المتجمع النازل . فمثلاً لو أردنا معرفة كم طالب يحصلون على 40

درجة فأكثر ، فالجواب يكون كل الطلبة أي 100 طالب ، وإذا أردنا معرفة كم طالب يحصل على 45 درجة فأكثر الجواب يكون (100 - 1 = 99 طالب) وبالمثل فعدد الطلبة الذين يحصلون على 50 درجة فأكثر يكون (99 - 4 = 95 طالب) وهكذا الى أن نصل الى 90 فأكثر فيكون عدد الطلبة صفر.

ولإيجاد التكرار المتجمع النازل نضيف عمودين لجدول التوزيع التكراري الذي يمثل درجات 100 طالب لمادة الاحصاء ، العمود الأول يعطي : الحد الأدنى للفئة فأكثر ، والعمود الثاني خاص بالتكرار المتجمع النازل : ثم بعد ذلك نبدأ بالفئة الأولى فأمام 40 فأكثر يكون التكرار المتجمع النازل 100 ، ثم بعد ذلك نحصل على التكرار المتجمع النازل للفئة الثانية عن طريق طرح تكرار الفئة من 100 وهكذا بالطرح المتتالي نحصل على التكرار المتجمع النازل كما هو موضح في الجدول أدناه . هذا ويمكن الحصول على نفس التكرار عن طريق الجمع المتتالي للتكرارات من أسفل الجدول . فأمام 90 فأكثر يكون التكرار المتجمع يساوي صفر ، وأمام 85 فأكثر يكون التكرار المتجمع (0 + 1 = 1) وأمام 80 فأكثر يكون التكرار المتجمع (1 + 3 = 4) وهكذا نحصل على التكرار المتجمع النازل عن طريق جمع التكرارات من أسفل .

جدول التوزيع التكراري والتكرار المتجمع النازل لدرجات
100 طالب في مادة الإحصاء

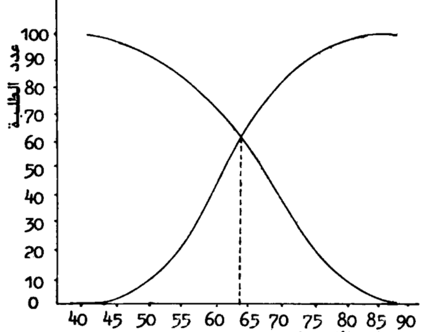
الفئات	التكرار	الحد الأدنى للفئة فأكثر	التكرار المتجمع النازل	التكرار المتجمع النازل النسبي
40	1	40 فأكثر	100	1.00
45	4	45 فأكثر	99	0.99
50	13	50 فأكثر	95	0.95
55	17	55 فأكثر	82	0.82
60	21	60 فأكثر	65	0.65
65	18	65 فأكثر	44	0.44
70	15	70 فأكثر	26	0.26
75	7	75 فأكثر	11	0.11
80	3	80 فأكثر	4	0.04
85وأقل من 90	1	85 فأكثر	1	.01
المجموع	100	90 فأكثر	0	0

ولإيجاد التكرار المتجمع النسبي نقسم التكرار المتجمع النازل على مجموع التكرارات كما في العامود الأخير من الجدول اعلاه .

ويمكن تمثيل التكرار المتجمع الصاعد بيانياً عن طريق المنحنى المتجمع الصاعد . ولرسم هذا المنحنى نأخذ الفئات على المحور الأفقي والتكرار المتجمع على المحور الرأسي . ثم نقوم بتعيين النقطة التي يكون إحداثيها السيني هو الحد الأعلى للفئة وإحداثيها الصادي هو التكرار المتجمع الصاعد. فالنقطة الأولى في مثالنا هذا تكون فوق 40 على ارتفاع صفر ، والنقطة الثانية تكون فوق 45 وعلى ارتفاع واحد . وهكذا حتى نصل الى النقطة الأخيرة ثم نوصل بين هذه النقط للحصول على المنحنى المتجمع الصاعد كما هو موضح في الشكل التالى.

ومن هذا المنحنى يمكننا إيجاد عدد الطلبة الذين حصلوا على درجات أقل من قيمة معينة ، فإذا أردنا معرفة عدد الطلبة الذين حصلوا على أقل من 70 مثلاً ، نرسم عموداً رأسياً فوق 70 على المحور الأفقي ، حتى يقابل هذا الخط المنحنى المتجمع الصاعد في نقطة معينة ، ثم نرسم خطاً أفقياً يقابل المحور الرأسي في نقطة معينة ، هذه النقطة هي التي تحدد عدد الطلبة الذين حصلوا على درجات أقل من 70 وعددهم 74 طالباً كما يمكن إيجاد نسبة للطلبة الذين حصلوا على درجات أقل من 70 عن طريق قسمة 74 على العدد الكلي وهو 100 في هذه الحالة تكون النسبة 74% .

المنحنى المتجمع الصاعد و المنحنى المتجمع النازل



ولرسم المنحنى المتجمع النازل نرسم المحورين ثم نحدد النقط فوق حدود الفئات كما هو الحال في رسم المنحنى المتجمع الصاعد . والشكل أعلاه يبين المنحنى المتجمع النازل على نفس الرسم مع المنحنى المتجمع الصاعد . ومن الجدير بالذكر أن هذين المنحنيين يلتقيان عند النقطة التي يكون احداثيها السيني هو الوسيط وهو أحد مقاييس النزعة المركزية على النحو المبين في الفصل التالي . ومن الرسم يمكننا الحصول على عدد الطلبة الذين تزيد درجاتهم عن قيمة معينة على النحو المبين بالنسبة للمنحنى الصاعد .

أمثلة تطبيقيـــة

مثـــال (الجمع بين فئتين أو أكثر)

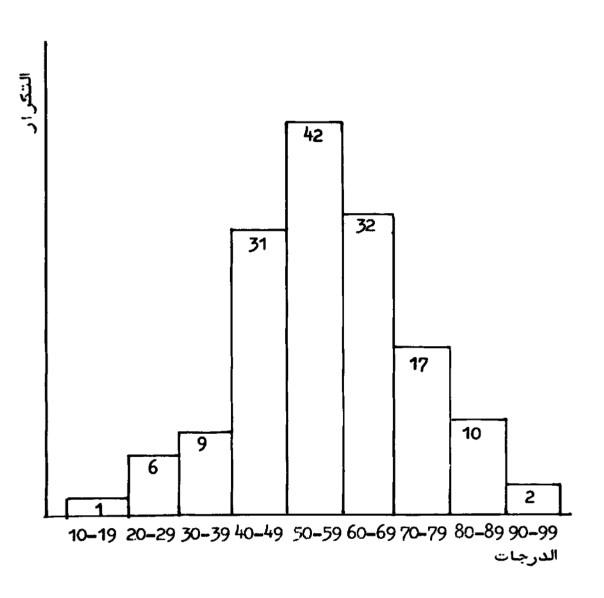
جدول التوزيع التكراري لاختبارات الذكاء

الدرجـــات	التكـــرارات
10 — 19	1
20 — 29	6
30 — 39	9
40 — 49	31
50 — 59	42
60 — 69	32
70 — 79	17
80 — 89	10
90 — 99	2
المجمـــوع	150

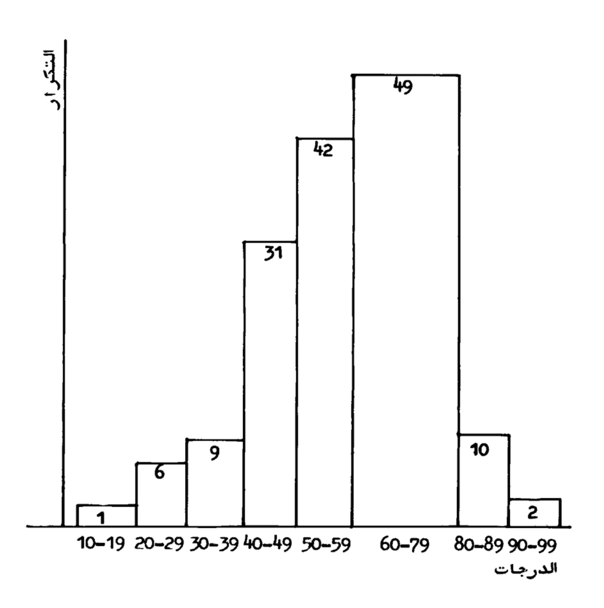
الجدول أعلاه يوضح تسعة فئات للدرجات التي حصل عليها الطـلاب في اختبار الذكاء كما أن الشكل (1) يوضح المدرج التكراري لهذه الدرجات .

لنفـرض أننا قمنـا بتجميع الفئـة 60 ـ 69 ، 70 ـ 79 ، والتي تقـابلهـا التكرارات ٣٢ ، 17 على التوالي . وبالحصول على الفئة الجديدة 60 ـ 79 نحصل على تكرار قدره 49 ، لكن إذا نظرنا الى الشكل (2) نجد أننا استعملنا ارتفاع المستطيل ليمثل تكرار الفئـة وهذا غـير صحيح . وفي هـذا المثال أطـوال الفئات متسـاوية ولكن الفئـة 60 ـ 79 في شكـل (2) هي عبـارة عن ضعف الفئـات الأخرى ، ولتعويض ذلك لابد من قسمة ارتفاع المستطيل على 2 كما هو موضح على شكل (3) الذي يمثّل المدرج التكراري المعدل (الصحيح) .

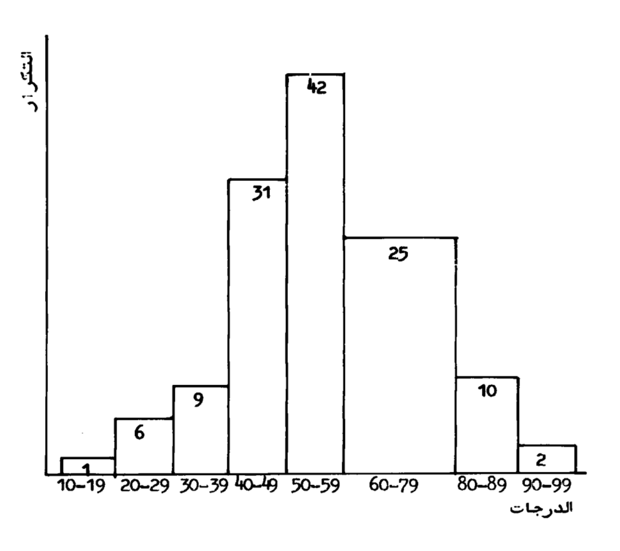
شكل (1) مدرج تكراري لاختبارات الذكاء



شكل (2) مدرج تكراري معدل لاختبارات الذكاء (غير صحيح)



شكل (3) مدرج تكراري معدل لاختبارات
الذكاء (صحيح)



تمارين

(1) أكتب الحدود لكل من رموز التجميع التالية :

(أ) $\sum_{j=1}^{n}(X_j+2)$

(ب) $\sum_{j=1}^{3}U_j(U_j+6)$

(جـ) $\sum_{j=1}^{4}4X_jY_j$

(د) $\sum_{j=1}^{4}f_jX^2_j$

(هـ) $\sum_{k=1}^{N}(Y^2_k-4)$

عبّر عما يلي باستخدام رموز التجميع :

(أ) $(X_1+3)^3 + (X_2+3)^3 + (X_3+3)^3$

(ب) $f_1(Y_1-a)^2 + f_2(Y_2-a)^2 + \ldots + f_{15}(Y_{15}-a)^2$

(جـ) $(2X_1-3Y_1) + (2X_2-3Y_2) + \ldots + (2X_n-3Y_n)$

(د) $(X_1 / Y_1-1)^2 + (X_2 / Y_2-1)^2 + \ldots + (X_8 / Y_8-1)^2$

(هـ) $\dfrac{f_1a^2_1 + f_2a^2_2 + \ldots + f_{12}a_{12}^2}{f_1 + f_2 + \ldots + f_{12}}$

(2) الجدول رقم (1) أدناه يبين الدخل الأسبوعي لموظفي إحدى المؤسسات

جدول رقم (1)

الدخل الأسبوعي (دينار)	عدد الموظفين
أقل من 90	43
90 - 94	155
95 - 99	620
100 - 104	1104
105 - 109	1423
110 - 114	1500
115 - 119	912
120 - 124	305
125 - 129	180
أكثر من 130	58
المجموع	6300

المطلوب حساب الآتي:

(أ) عدد الموظفين الذين يتقاضون مرتباً لا يقل عن 104 ديناراً في الأسبوع؟

(ب) عدد الموظفين الذين يتقاضون مرتباً يزيد عن 114 ديناراً في الأسبوع؟

(جـ) عدد الموظفين الذين يتقاضون مرتباً يزيد عن 150 ديناراً في الأسبوع؟

(د) عدد الموظفين الذين يتقاضون مرتباً أقل من 120 ديناراً في الأسبوع؟

(هـ) عدد الموظفين الذين يتقاضون مرتباً وقدره 109 ديناراً على الأكثر؟

(و) عدد الموظفين الذين يتقاضون مرتباً وقدره 129 ديناراً أو أقل؟

(3) أجرت إدارة اطفاء الحريق في احدى الدول معاينة مع 435 متقدماً للإلتحاق بخدمات الحريق وقامت بأخذ أوزانهم لأقرب $\frac{1}{10}$ من الرطل وكان أقل وزن للمتقدمين 153.2 رطلاً كما كان أكبر وزن 223.7 رطلاً.

المطلــوب :

عمل ثلاثة جداول تكرارية يحتوي كل منها على ثماني فئات متساوية الطول على أن يوضح الجدول الأول مراكز الفئات والجدول الثاني الحد الأدنى والأعلى للفئات والجدول الثالث يوضح الحد الأدنى والأعلى الحقيقي للفئات.

(4) الأرقام أدناه توضح عدد سيارات الشحن التي غادرت أحد الموانىء في خلال 120 يوماً .

59	46	64	63	53	71	41	60	51	55	64	50
55	50	61	55	65	58	62	65	57	61	45	66
60	64	49	59	46	64	56	53	66	58	57	53
51	57	68	61	63	56	62	59	47	42	64	58
62	60	43	64	58	52	52	67	59	60	51	61
52	54	66	53	69	60	73	55	63	56	63	48
57	62	56	62	56	63	59	61	60	65	59	56
47	58	52	67	43	59	64	58	48	63	52	57
58	63	66	60	78	65	61	57	67	54	53	63
54	61	55	65	63	49	62	84	59	61	55	60

المطلوب :

(أ) تجميع هذه البيانات في جدول توزيع تكراري محتوياً على الفئات التالية : ـ

40 — 44, 45 — 49, 50 — 54, 55 — 59,

60 — 64, 65 — 69, 70 — 74, 75 — 79

(ب) رسم مدرج تكراري من الجدول الذي تحصلت عليه في (أ)

(جـ) حول جدول تكرار الفئات الى نسب مئوية ثم أحصل على المضلع التكراري .

(د) حول النسب المئوية في (جـ) الى نسب مئوية تجميعية في صيغة (أقل من) ثم أرسم التكرار المتجمع الصاعد.

(٥) الخبر التالي أوردته صحيفة الوطن الكويتية في عددها الصادر يوم الأربعاء الموافق 16/9/1981 .

المطلوب دراسة هذا الخبر وإجراء الآتي:

أ - 1 - عمل جدول تكراري لأعمار الأزواج .

2 - رسم مدرج تكراري للجدول في (1) .

3 - عمل جدول تكراري لأعمار الزوجات .

4 - رسم مضلع تكراري لأعمار الزوجات .

ب - 1 - عمل جدول تكراري لأعمار الأزواج المطلقين .

2 - عمل جدول تكراري لأعمار الزوجات المطلقات .

جـ - عمل جدول تكراري لعدد سنين الزواج .

***

« 630 زواجاً و 151 حالة طلاق خلال الشهرين الماضيين »

بلغ إجمالي عقود الزواج المبرمة خلال شهري يوليو وأغسطس 630 عقداً كان مجموع العقود منها خلال شهر أغسطس 314 منها 66 عقداً جنسية الزوج فيها من الكويتيين و248 من غير الكويتيين في حين بلغ مجموع الزوجات الكويتيات 70 زوجة و244 زوجة من غير الكويتيات.

أما بالنسبة لأعمار الأزواج الذين عقدوا زواجهم خلال شهر أغسطس في سن تحت 19 سنة كان عددهم 15 في حين تزوج 113 شخصاً في سن من 24—20 سنة وأبرم 101 شخصاً عقود زواجهم في سن من 29—25 سنة ، أما من تزوجوا في سن 34—30 سنة فكان مجموعهم 4؛ زوجاً وتزوج 30 شخصاً ممن بلغوا سن 44—35 سنة وتزوج 14 شخصاً تجاوزت أعمارهم سن 45 سنة .

أما بالنسبة لأعمار الزوجات اللاتي عقد قرانهن تحت سن 15 سنة فكان مجموعهن 5 زوجات وتزوجت 137 فتاة ممن بلغت أعمارهن من 19— 15 سنة في حين أبرمت عقدو زواج 99 فتاة ممن تراوحت أعمارهن من 24—20 سنة ، أما من بلغت أعمارهن من 29—25 سنة فكان مجموعهن 42 زوجة وتزوجت 11 امرأة ممن بلغت أعمارهن 34—30 سنة كما تزوجت 15 امرأة أخرى ممن تراوحت أعمارهن ما بين 40—35 سنة وعقد قران 5 زوجات أعمارهن فوق سن 45 سنة .

أما بالنسبة لحالات الطلاق التي وقعت خلال شهر أغسطس الماضي فبلغ مجموعها 151 طلاقاً فقد بلغ مجموع حالات الطلاق في شهر يوليو الماضي 125 حالة أي بفارق 26 طلاقاً بين الشهرين كان مجموع المطلقين من الكويتيين 97 شخصاً في حين كان مجموع المطلقين من غير الكويتيين 54 شخصاً و75 مطلقة كانت من الكويتيات في حين بلغ مجموع المطلقات من غير الكويتيات 76 مطلقة .

وكانت فئات أعمار ازواج الذين أنهوا عقود الزواج ممن تتراوح أعمارهم من 19—15سنة زوج واحد و26 زوجاً فسخوا عقد زواجهم في سن يتراوح من24—20 و37 زوجاً تراوحت أعمارهم من 29—25 سنة و24 زوجاً تراوحت أعمارهم من 34—30 سنة أما الذين طلقوا زوجاتهم وهم في سن من 44—35 سنة 35 زوجاً و27 زوجاً فسخوا زواجهم وهم في سن 45 سنة وكان عدد المطلقات في سن بين 15 ـ 19 سنة 25 زوجة وطلقت 52 زوجة وهن في سن 24—20 سنة و24 مطلقة فسخت عقود زواجهن وهن في سن يتراوح بين 39—25 سنة وكان عدد من طلقن

وهن في سن يتراوح بين 30 ـ 34 سنة 23 مطلقة و16 مطلقة وهن في سن يتراوح من 35 ـ 40 سنة و10 زوجات فسخن عقود زواجهن وهن في عمر 45 سنة وقد تراوحت مدد الحياة الزوجية التي استمرت بين الزوجين وانتهت بوقوع الطلاق على النحو التالي:

59 حالة طلاق وقعت ولم تستمر الحياة الزوجية أكثر من سنة في حين استمر 34 زواجاً لمدة تراوحت بين سنة ـ 5 سنوات انتهت بالطلاق ، أما بالنسبة لمن فسخوا عقود زواجهم ممن استمرت حياتهم الزوجية من 5 ـ 10 سنوات فبلغ عددهم 31 طلاقاً و17 طلاقاً وقع بعد مضي 10 ـ 20 سنة على الحياة الزوجية وفسخت عقود 7 أزواج ممن أمضوا 20 سنة في حياة زوجية مشتركة.

الفصـــل الثاني

مقاييس النزعة المركزية

MEASURES OF CENTRAL TENDENCY

الفصـــل الثاني

مقاييس النزعة المركزية

MEASURES OF CENTRAL TENDENCY

ما عرضنا له في الفصـل الأول خاص بتحليل البيانـات من خلال أسلـوب العرض البياني ، هذا الأسلوب سهل وبسيط ويمتاز بسرعة الإقناع ، يسـتقيم مع العديد من الباحثين الـذين لا يميلون الى استخدام أسـاليب أخـرى في الدراسـة والتحليـل ويعتمد هذا الأسلوب على نحو ما أسلفنا على تمثيل الجداول الإحصائية بيانياً من خلال منحنـى الظاهـرة ، حيث يمـكن تحـديد العلاقــات والخصــائص والإتجاهات على أساسه.

غير أن الأسلوب البيانـي في تحليل ودراسـة الظواهـر لتحـديد الخصـائص والإتجاهات والعلاقات ، يعتمد في دقته على دقة التمثيل البياني نفسه وبذلك ربما تختلف الخصائص من رسم الى آخر لنفس الظاهرة ، كما أنه قد تتشابه المنحنيات في الأشكال المختلفة للتوزيعات التكرارية ، وبالتالي يصعب استخدام هذا الأسلوب كأساس لتحديد الخصائص والإتجاهات وللمقارنات بين التوريعات التكرارية والتي تنتمي الى مجموعة تكرارية واحدة.

لهذه الأسباب فإنه من الأفضل اللجوء الى أساليب القياس الكمية ، حيث يستخدم الباحث الطريقة الرياضية في قياس التغـيرات وتحـديد درجـة العلاقـات وطبيعة الإتجاهات بطريقة موضوعية غير متحيزة.

في هذا الفصل نعرض لأحد المجموعات الـرئيسية في طرق القياس الكمـي

لقيم التمركز في الظواهر وذلك لتحديد الإتجاهات العامة وتلخيص البيانات من خلال هذه القيم .

المتوسطات (مقاييس النزعة المركزية) :

المتوسط هو القيمة النموذجية أو الممثلة لمجموعة من البيانات ـ وحيث أن مثل هذه القيمة النموذجية تميل الى الوقوع في المركز داخل مجموعة بيانات مرتبة حسب قيمتها ، فإن المتوسطات تسمى أيضاً بمقاييس النزعة المركزية . ويمكن أن نعرف صوراً عديدة للمتوسطات وإن كان الأكثر شيوعاً الوسط الحسابي أو باختصار الوسط ، الوسيط ، المنوال ، الوسط الهندسي والوسط التوافقي ـ وكل منها له مميزاته وعيوبه وهذا يعتمد على البيانات والهدف من استخدامه .

الوسط الحسابي* : The Arithmatic Mean

الوسط الحسابي أو الوسط للمجموعة n من الأرقام $X_1, X_2, X_3, \dots, X_n$ ويرمز له بالرمز $\bar{X}$ (ويقرأ «X bar») ويعرف كالآتي :

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} = \frac{\sum_{j=1}^{n} X_j}{n} = \frac{\sum X_j}{n} \qquad (1)$$

مثال : الوسط الحسابي للأرقام 8, 3, 5, 12, 10 هو :

$$\bar{X} = \frac{8+3+5+12+10}{5} = \frac{38}{5} = 7.6$$

إذا كانت الأرقام $X_1, X_2, \dots, X_k$ تحدث $f_1, f_2, \dots, f_k$ مرة على الترتيب بمعنى

---

* ملحوظة :

يعرف الوسط الحسابي للمجتمع بأنه مجموع مشاهدات المجتمع مقسومة على عدد مشاهدات المجتمع N ويرمز للوسط الحسابي بالحرف μ وعليه فإن :

$$\mu = \frac{X_1 + X_2 + \dots + X_N}{N}$$

وإذا كانت مشاهدات المجتمع تنقسم الى k مجموعة وتكرار المجموعات $f_1, f_2, \dots f_k$

$$\mu = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{\sum f_i X_i}{N}$$

أنها تحدث بتكرارات $(f_1, f_2, ..., f_k)$ فإن الوسط الحسابي سيكــون

$$(2) \qquad \bar{X} = \frac{f_1X_1 + f_2X_2 + ... + f_k X_k}{f_1 + f_2 + ... + f_k} = \frac{\sum_{j=1}^{k} f_j X_j}{\sum_{j=1}^{k} f_j} = \frac{\Sigma fX}{\Sigma f} = \frac{\Sigma fx}{n}$$

حيث $n = \Sigma f$ وهو مجموع التكرارات أي مجموع عدد الحالات.

مثال: إذا كانت 5, 8, 6, 2 تحدث بتكرارات 3, 2, 4, 1 على الترتيب فإن الوسط الحسابي سيكـون :

$$\bar{X} = \frac{(3)(5) + (2)(8) + (4)(6) + (1)(2)}{3 + 2 + 4 + 1} = \frac{15 + 16 + 24 + 2}{10} = \frac{57}{10}$$

$$= 5.7$$

الوسط الحسابي المرجــح : The Weighted Mean

في بعض الأحيان نقرن بعض الأرقام $X_1, X_2, ..., X_k$ بمعامـلات ترجيح أو أوزان $W_1, W_2, ..., W_k$ وهذه تعتمد على الدلالة أو الأهمية المرتبطة بهذه الأرقام في هذه الحالة

$$(3) \qquad \bar{X} = \frac{W_1X_1 + W_2X_2 + ... + W_kX_k}{W_1 + W_2 + ... + W_k} = \frac{\Sigma W_iX_i}{\Sigma W_i} = \frac{\Sigma WX}{\Sigma W}$$

يسمى بالوسط الحسابي المرجح ، لاحظ أوجه الشبه بالمعادلة (2) التي يمكن اعتبارها وسطاً حسابياً مرجحاً بأوزان $f_1, f_2, ..., f_k$

مثـال: إذا كان الامتحان النهائي في مقرر أعطى وزناً ثلاثة أمثال الامتحانات الشفهية وإذا حصل طالب في الإمتحان النهائي على 85 وفي الإمتحانات الشفهية على 70, 90 فإن متوسطة تقديــره هو

$$\bar{X} = \frac{(1)(70) + (1)(90) + (3)(85)}{1 + 1 + 3}$$

$$= \frac{70 + 90 + 255}{5} = \frac{415}{5} = 83$$

خصائص الوسط الحسـابي :

(أ) المجموع الجبري لانحرافات إلقيم عن وسطها الحسابي يساوي صفراً .

$$\sum_{i=1}^{n} (X_i - \overline{X}) = 0$$

مثـال: انحرافات الأرقام 10 ,12 ,5 ,3 ,8 عن وسطها الحسابي 7.6 هو

$$5 - 7.6 = -2.6, 3 - 7.6 = -4.6, 8 - 7.8 = 0.4,$$

$$12 - 7.6 = 4.4, 10 - 7.6 = 2.4$$

وعليه فإن

$$\Sigma (X_i - \overline{X}) = -2.6 - 4.6 + 0.4 + 4.4 + 2.4$$

$$= -7.2 + 7.2 = 0$$

(ب) مجموع مربعات انحرافات مجموعة من الأرقام X_i عن أي رقم a يكون أصغر ما يمكن في حالة واحدة فقط إذا كانت $a = \overline{X}$

(جـ) الوسط الحسابي لعدة مجموعات عبارة عن الوسط الحسابي المرجح لكل وسط حسابي لكل مجموعة مرجحاً بحجم هذه المجموعة .

$$\overline{X} = \frac{f_1 m_1 + f_2 m_2 + \dots + f_k m_k}{f_1 + f_2 + \dots + f_k} \qquad (4)$$

حيث m_i هي الوسط الحسابي للمجموعة i , f_i عدد أفراد هذه المجموعة .

مثال : أخـذت عينة عشوائية من خمسين عاملاً من عمال أحـد المصانـع الحربية فوجد أن متوسط أجر العامل 75 ديناراً شهرياً ومن عينة أخرى من مائة عامل من عمال أحد مصانع المعلبات فوجد أن متوسط أجـر العامـل 49.2 ديناراً ومن عينة ثالثة لأحد مصانع الحديد والصلب من مائة وخمسين عامـلاً وجد أن متوسط أجر العامل الشهري 105 · ديناراً .

المطلوب إيجاد الوسط الحسابي لأجر العامل الشهري في المصانع الثلاثة .

$$\overline{X} = \frac{50\times 75 + 100\times 49.2 + 150\times 105}{50 + 100 + 150}$$

$$\overline{X} = \frac{3750 + 4920 + 15750}{300} = \frac{24420}{300} = 81.400$$

(د) إذا كانت A أي وسط حسابي افتراضي أو مخمن (والذي يمكن أن يكون أي رقم) وإذا كان $d_j = X_j - A$ هو انحرافات X_j عن A فإن المعادلات السابقة (1) ، (2) سيصبحان على الترتيب

$$(5) \qquad \overline{X} = A + \frac{\sum_{j=1}^{n} d_j}{n} = A + \frac{\Sigma d}{n}$$

$$(6) \qquad \overline{X} = A + \frac{\sum_{j=1}^{n} f_j d_j}{\Sigma f_j} = A + \frac{\Sigma fd}{n}$$

حيث $n = \sum_{j=1}^{n} f_j = \Sigma f$ لاحظ أن (5) و(6) يمكن تلخيصهما بالمعادلة

$$\overline{X} = A + \overline{d}$$

الوسط الحسابي محسوباً من بيانات مجمعــة :

عندما تعرض البيانات في توزيع تكراري ، فإن جميع القيم التي تقع داخل فئة معينة تعتبر أنها مطابقة لمركز الفئة أو منتصف مدى الفئة. الصيغتان (2) و (6) يمكن استخدامهما للبيانات المجمعة إذا اعتبرنا X_j مركز الفئة و f_j التكرار المقابل لها و A أي مركز فئة افتراضي أو مخمن و $d_j = X_j - A$ انحرافات X_j عن A . والحساب بالطريقة (2) ، (6) ، أي أن $\overline{X} = A + \frac{\Sigma f_i d_i}{\Sigma f_i}$ ، $\overline{X} = \frac{\Sigma f_i X_j}{\Sigma f_i}$ ، يسميان أحياناً بالطريقة المطولة والطريقة المختصرة على الترتيب.

وإذا كانت أطوال الفئات متساوية وتساوي C ، والإنحرافات $d_j = X_j - A$ يمكن التعبير عنها بالصورة Cu_j حيث u_j يمكن أن يكون عدداً صحيحاً موجباً أو سالباً أو صفراً ، أي . . . $0, \pm 1, \pm 3,$ وبهذا فإن الصيغة (6) تصبح

$$\bar{X} = A + \left(\frac{\sum_{j=1}^{k} fu_j}{n}\right) C = A + \left(\frac{\Sigma fu}{n}\right) C \qquad (7)$$

وهذه الطريقة تسمى بطريقة الترميز عند حساب الوسط الحسابي ، وهي طريقة مختصرا جدا وتستخدم في حالة البيانات المجمعة عندما تكون أطول الفئات متساوية . لاحظ أنه في طريقه الترميز نجد قيم المتغير X تحول إلى قيم المتغير u بالعلاقة $X = A + Cu$

مثـــال :

الجدول يوضح أوزان 100 طالب في إحدى الجامعات

عدد الطلبة	الأوزان كيلوجرامات
5	60- 62
8	63- 65
42	66- 68
27	69- 71
8	72- 74
100	**المجمـــوع**

المطلوب إيجاد الوسط الحسابي لأوزان المائةطالب في الجامعة باستخـدام (أ) الطريقة المختصرة (ب) طريقة الترميز.

الحـــل :

(أ) يمكن أن ينظم الحل كما في الجدول أدناه . دعنا نأخذ A مركز الفئة الثالثة أي67 (المقابل لأكبر تكرار)، على الرغم من أن أي مركزفئة يمكن استخدامه.

الفئات	مراكز الفئات X_i	انحرافات $d_i = X_i - A$	تكرارات f_i	$f_i d_i$
60 - 62	61	−6	5	− 30
63 - 65	64	−3	18	− 54
66 - 68	A = 67	0	42	0
69 - 71	70	3	27	81
72 - 74	73	6	8	48
المجموع			$n = \sum_{i=1}^{k} f_i = 100$	$\sum f_i a_i = 45$

(أ) باستخدام الطريقة المختصرة

$$\overline{X} = A + \frac{\sum_{i=1}^{k} f_i d_i}{\sum_{i=1}^{k} f_i} = A + \frac{\sum fd}{n}$$

حيث

$$\sum_{i=1}^{k} f_i = \sum f = n$$

إذا

$$\overline{X} = 67 + \frac{45}{100} = 67.45 \text{ Kg}$$

(ب) باستخدام طريقة الترميز يمكن ترتيب الحل كما في الجدول الآتي:

X	u_i *	f_i	f_iu_i
61	− 2	5	− 10
64	− 1	18	− 18
A 67	0	42	0
70	1	27	27
73	2	8	16
	* $u_i = \frac{X_i - 67}{3}$	n = 100	Σ fu = 15

وبتطبيق صيغة الترميز نحصــل على

$$X = A + \frac{(\Sigma fu)}{n} C = 67 + (\frac{15}{100})(3) = 67.45$$

الوسيط: The Median

الوسيط لمجموعة من الأرقام مرتبة حسب قيمها (في منظومة) هي القيمة التي في المنتصف أو الوسط الحسابي للقيمتين بالمنتصف.

مثال 1 ـ مجموعة الأرقام 3, 4, 4, 5, 6, 8, 8, 8, 10 وسيطها هو6

مثال 2 ـ مجموعة الأرقام5, 5, 7, 9, 11, 12, 15, 18 وسيطها هو10 = $\frac{1}{2}(9 + 11)$

أما في البيانـات المجمعـة فإن الـوسيط نحصـل عليه بالإستكمال ويحسب كالآتي:

$$\text{الوسيط} = \tilde{X} = L_1 + \left(\frac{\frac{n}{2} - (\Sigma f)_1}{f_{median}}\right) C \qquad (8)$$

حيث:

L_1 = الحد الأدنى الحقيقي للفئة الوسيطية (أي الفئة التي يقع فيها الوسيط).

n = عدد العناصر في البيانات (مجموع التكرارات).

$(\Sigma f)_1$ = مجموع التكرارات لجميع الفئات قبل الفئة الوسيطية.

f_{median} = تكرار الفئة الوسيطية.

C = طول الفئة الوسيطية.

ويمكن التعبير هندسياً عن الوسيط بأنه القيمة X على الإحداثي السيني التي إذا رسم عندها عمود رأسي فإنه يقسم المدرج التكراري الى جزءين متساويين . يعبر عن هذه القيمة أحياناً بـ $\tilde{X}$

مثـــال :

(١) الأجر بالساعة لخمسة عاملين في مصنع هو $2.52, $3.28, $3.96, $3.75, $9.20 أوجد.

(أ)وسيط أجر الساعة . (ب) الوسط الحسابي لأجر الساعة

الحـــل:

(أ) بترتيب الأجور في منظومة تصبح كالآتي:

$2.52, $3.28, $3.75, $3.96, $9.20

وبما أن هناك عدداً فردياً من القيم فإن هناك قيمة واحدة في المنتصف وهـي 3.75 $ وهي الوسيط المطلـــوب .

(ب) الوسط الحسابي هو:

$$\frac{2.52 + 3.96 + 3.28 + 9.20 + 3.75}{5} = \$4.54$$

لاحظ أن الـوسيط لم يتأثـر بالقيمـة المتطرفـة 9.20 $ بينما تأثـر بهـا الوسـط الحسابي . وفي هذه الحالة فإن الوسيط يعطي دلالة أفضل على معدل أجر الساعة عن الوسط.

(2) الجدول الآتي يوضح أطوال 40 من الطلاب أوجد الطول الوسيـــط؟

الطــو ل	التكــرار
118 - 126	3
127 - 135	5
136 - 144	9
145 - 153	12
154 - 162	5
163 - 171	4
172 - 180	2
المجمـوع	40

يمكن الحل بطريقتين .

الطريقة الاولى :

في هذه الحالة فإن الوسيط هو الطول الذي يقع نصف التكرار الكـلي قبله ($\frac{40}{2} = 20$) والنصف الآخر بعده.

وحيث أن مجموع تكرارات الفئات الثلاثة الأولى هو $17 = 9 + 5 + 3$ وحتى نحصل على الرقم المطلوب 20 فإننا نحتاج إلى 3 أرقام من الـ 12 حالة الموجودة في الفئة الرابعة .

وبما أن الفئة الرابعة 153 - 145 هي في الحقيقة تقابـل الأطـوال144.5 — 153.5 فإن الوسيط يقع في $\frac{3}{12}$ المسافة بين 144.5 و 153.5 أي أن الوسيط هو

$$144.5 + \frac{3}{12}(153.5 - 144.5) = 144.5 + \frac{3}{12}(9) = 146.8\text{Cm}$$

أما الطريقة الثانية فباستخدام القانون حسب الصيغة (7)

بما أن مجموع التكرارات المقابلة للفئات الثلاث الأولى والفئات الأربع الأولى على الترتيب $3 + 5 + 9 = 17$ ، $3 + 5 + 9 + 12 = 29$ فإن الوسيط يقع في الفئة الرابعة والتي هي بالتالي الفئة الوسيطية . وبهذا :

L_1 = الحد الأدنى للفئة الوسيطية . = 144.5

n = عدد العناصر في البيانات . = 40

$(\Sigma f)_1$ =مجموع التكرارات لجميع الفئات قبل الفئة الوسيطية

$$3 + 5 + 9 = 17$$

f_{median} = تكرار الفئة الوسيطيـة وهـو 12

C = طول الفئة الوسيطيـة وهـو 9

وبهـذا فـإن

$$\tilde{X} = \text{الوسيط} = L_1 + \left(\frac{\frac{n}{2} - (\Sigma f)_1}{f_{median}} \right) C$$

$$= 144.5 + \left(\frac{\frac{40}{2} - 17}{12} \right)(9) = 146.8\,Cm$$

المنـوال : The Mode

المنوال لمجموعة من القيم هي القيمة التي تتكرر أكثر من غيرها أو القيمة الأكثر شيوعاً . وقد لا يكون للقيم منوال وقد يوجد منوال ولكنه غير وحيد .

مثـال 1 ـ المجموعة 22, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 لها منوال وهو9

مثـال 2 ـ المجموعة 3, 5, 8, 10, 15, 16 ليس لها منوال

مثال 3 ـ المجموعة 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9 لها منوالان وهما 4 ,7 وتسمـى

مجموعة ذات منوالين Bimodal

التوزيع الذي له منوال واحد يسمى وحيد المنوال Unimodal

في حالة البيانات المجمعة حيث يعبر عن البيانات بمنحنى تكراري فإن المنوال هو قيمة (أو قيم : X المقابلة لنقطة (أو نقط) النهاية العظمى للمنحنى ويعبر أحياناً عن هذه القيمة لـ X بالرمز $\widehat{X}$ ونحصل على المنوال من التوزيع التكراري أو المدرج التكراري بالصيغة :

$$\widehat{X} = \text{المنوال} = L_1 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) C \qquad (9)$$

حيث :

L_1 = الحد الأدنى الحقيقي للفئة المنوالية (أي الفئة التي يقع فيها المنوال)

Δ_1 = زيادة تكرار الفئة المنوالية عن تكرار الفئة قبل المنوالية .

Δ_2 = زيادة تكرار الفئة المنوالية عن تكرار الفئة بعد المنوالية .

C = طول الفئة المنوالية .

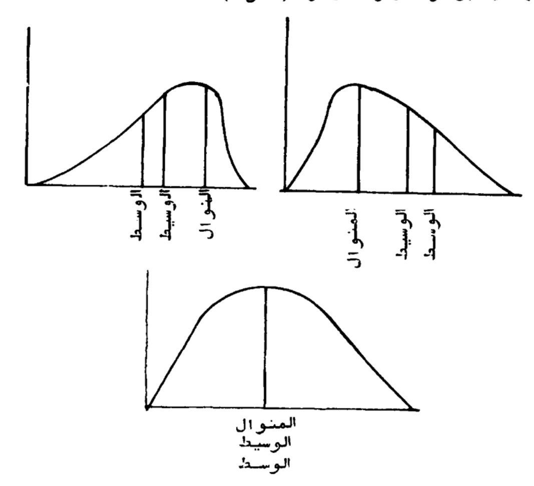
علاقة اعتبارية بين الوسط والوسيط والمنوال :

المنحنيات التكرارية وحيدة المنوال والبسيطة الإلتواء (غير متماثلة) تحقق العلاقة الاعتبارية :-

الوسط - المنوال = ٣ (الوسط - الوسيط)

$$\overline{X} - \widehat{X} = 3(\overline{X} - \widetilde{X}) \qquad (10)$$

الأشكال (1) ، (2) أدناه توضح الموضع النسبي للوسيط والوسيط والمنـوال للمنحنيات التكرارية الملتوية الى اليمين والمنحنيات الملتوية لليسار أما في المنحنيات المتماثلة يتطابق الوسط والوسيط والمنوال (شكل 3).



الوسط الهندسي : Geometric Mean

ويرمز له بالرمز G والوسط الهندسي لمجموعة n من الأرقام $X_1, X_2, X_3, ..., ..., X_n$ هو الجزر النوني لحاصل ضرب هذه الأرقام

$$G = \sqrt[n]{X_1, X_2, X_3, ..., X_n} \quad (11)$$

مثال: الوسط الهندسي للأرقام 2, 4, 8 هو

$$G = \sqrt[3]{(2)\ (4)\ (8)} = \sqrt[3]{64} = 4$$

ومن الناحية العملية فإن الوسط الهندسي G يحسب عادة باستخدام اللوغريثمات .

الوسط التوافقي : Harmonic Mean

ويرمز له بالرمز H والوسط التوافقي H لمجموعة من n من الأرقام $X_1, X_2, X_3, \ldots, X_n$ هو مقلوب الوسط الحسابي لمقلوبات هذه القيم .

$$(12) \qquad H = \frac{1}{\frac{1}{n}\sum_{j=1}^{n}\frac{1}{X_j}} = \frac{n}{\Sigma \frac{1}{X}}$$

ومن الناحية العملية فقد يكون من الأسهل أن نتذكر أن

$$(13) \qquad \frac{1}{H} = \frac{\Sigma \frac{1}{X}}{n} = \frac{1}{n}\ \Sigma \frac{1}{X}$$

مثـال: الوسط التوافقي للأرقام 2.4, 8 هو

$$H = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}} \quad \frac{3}{\frac{7}{8}} = 3.43$$

العلاقة بين الوسط الحسابي والوسط الهندسي والوسط التوافقي :

الوسط الهندسي لمجموعة من الأرقام $X_1, X_2. \ldots, X_n$ أقـل من أو يساوي وسطها الحسابي ولكنه أكبر من أو يساوي وسطها التوافقي.

$$H \leqslant G \leqslant \bar{X}$$

و تتحقق علامة التساوي إذا كانت الأرقام $X, X_2, \ldots, X_n$ متساوية .

مثال: المجموعة 2, 4, 8 وسطها الحسابي 4.67 ووسطها الهندسي 4 ووسطها التوافقي 3

الربيعات والعشيرات والميئنات Quartiles, Deciles, Percentiles

إذا رتبت مجموعة من الأرقام حسب قيمها فإن القيمة التي في المنتصف (أو الوسط الحسابي للقيمتين بالمنتصف) والتي تقسم مجموعة القيم إلى مجموعتين متساويتين في العدد هي الوسيط . وبتعميم هذه الفكرة يمكن أن نفكر في القيم التي تقسم المجموعة إلى أربعة أجزاء متساوية . هذه القيم يرمز لها بالرموز Q_1, Q_2, Q_3 تسمى بالربيع الأول ، الربيع الثاني ، الربيع الثالث على الترتيب ، القيمة Q_2 تساوي الوسيط ، كذلك فإن القيم التي تقسم المجموعة إلى عشرة أجزاء متساوية تسمى بالعشيرات فيرمز لها بالرمز $D_1, D_2, D_3, ..., D_9$ بينما القيم التي تقسم البيانات إلى مائة قسم متساو تسمى بالمئينات ويرمز لها بالرمز $P_1, P_2, P_3, ..., P_{99}$ العشير الخامس والمئين الخمسون يساويان الوسيط . المئين الخامس والعشرون والمئين الخامس والسبعون يساويان الربيع الأول والربيع الثالث على الترتيب . وإجمالاً يمكن إيجاد الربيعات والعشيرات والمئينات وغيرها من القيم بتقسيم البيانات إلى أقسام جزئية متساوية تسمى قيم التقسيمات الجزئية .

تمارين :

(1) لدينا المعلومات المكتملة عن مجموع الضرائب المتحصلة من 153 محلاً تجارياً في إحدى المدن وضح بالأمثلة متى تمثل هذه المعلومات

(أ) مجتمعاً (ب) عينة

(2) أحسب الوسط الحسابي والوسيط والمنوال للنسب المئوية للأرباح التي تتقاضاها عشر شركات استثمارية إذا كانت هذه النسب هي 9.5 9.5 8.75

9.5 8.5 9.25 9 8.5 10 9.5

(3) إذا كانت الحمولة القصوى لإحدى الشاحنات هي أربعة ألف رطل وضّح ما إذا كانت الحمولات الآتية تفوق هذا الحد الأقصى:

(أ) 25 صندوقاً متوسط وزنها 167 رطلاً للصندوق.

(ب) 10 صناديق صغيرة متوسط وزنها 65 رطلاً و20 صندوقاً كبيراً متوسط وزنها 150 رطلاً.

(جـ) 12 صندوقاً متوسطة الوزن، متوسط وزنها 115 رطلاً و15 صندوقاً كبيراً متوسط وزنها 180 رطلاً.

(4) تقوم إحدى المؤسسات الصناعية بمتابعة حسابات 15 من عملائها وذلك عن طريق حصر الديون المعلقة كنسبة مئوية من جملة مبيعاتها فإذا كانت هذه النسب في أحد الشهور كالآتي: -

1.0% 0.98% 0.99% 1.0% 1.1% 0.94% 0.80%

0.88% 0.90% 0.90% 1.01% 0.98% 0.95% 4.2%

0.80% 0.75%

مطلوب حساب الآتي:

(أ) الوسط الحسابي (ب) الوسيط (جـ) المنوال

لهذه النسب مع توضيح ملاءمة كل من هذه المقاييس لاعطاء صورة صحيحة عن وضع هؤلاء المدينين.

(5) تقدمت إحدى الهيئات بمشروع قرار لأحد المجالس التشريعية في إحدى الدول وذلك لالغاء الضريبة المفروضة على الأدوية. هذا وقد كان تعليق السيد وزير المالية في تلك الدولة كالآتي: -

«لقد كان متوسط ما دفعه الفرد الواحد خلال السنين الثلاثة الماضية دينارين فقط في العام وهذا المبلغ من الضآلة بمكان حيث أنه لا يمثل أي عبء على المواطن، لذلك فإنني لا أرى داعيا لالغاء الضريبة».

علق تعليقاً إحصائياً على مقولة السيد الوزير إذا كنت أحد الذين يدافعون عن إلغاء الضريبة .

(6) (أ) أوجد الوسط التوافقي H للأرقام التالية :

3 ، 5 ، 6 ، 6 ، 7 ، 10 ، 12

(ب) أوجد الوسط الهندسي للأرقام 4 ، 7 ، 8 ، 3 ، 5 ، 3 ، 2 .

(جـ) من الأرقام 6 ، 4 ، 2 ، صفر .

أوجد

(1) الوسط الحسابي .

(2) الوسط الهندسي .

(3) الوسط التوافقي .

الفصـــل الثالث

مقاييس التشتت

MEASURES OF DISPERSION

الفصــل الثالث

مقاييس التشتت

MEASURES OF DISPERSION

لا تعتبر مقاييس التمركز كمقياس كاف لوصف مجموعة من البيانات وصفـاً كاملاً فقد تتساوى بعض العينات في الوسط الحسابي بالرغم من اختلافها تماماً عن بعضها .فالعينات الثلاثة التالية ذات وسط حسابي واحد ولكنها بلا شك تختلف عن بعضها ، فأفراد العينة الأولى متساوية ويمثلها الوسط الحسابي 8 كذلك أفراد العينة الثانية يمثلها الوسط الحسابي 8 ولكن بعض الأفراد أقل والبعض الآخر أكبر من الوسط الحسابي ، أما العينة الثالثة فإن القيم أكثر اختلافاً من بعضها ولكن لها نفس الوسط الحسابي للعينات (1) و(2) .

	(1)	(2)	(3)
	8	7	4
	8	5	3
	8	8	6
	8	12	16
			11
الوسط الحسابي	8	8	8

فالوسط الحسابي يمثل مركز البيانات لكنه لا يبين مدى التفاف أو بعثرة

البيانات حول هذا الوسط ولهذا لا بد من وجود مقياس آخر مع المقاييس المركزية لقياس الإختلافات أو البعثرة في داخل العينات لمعرفة مدى قدرة المقياس المركزي على تمثيل البيانات . وعلى هذا لوصف أي مجموعة من البيانات يجب حساب أحد المقاييس المركزية مقروناً بأحد مقاييس التشتت .

الدرجة التي تتجه بها البيانات الرقمية للإنتشار حول قيمة وسطى تسمى تشتت أو تغير البيانات . وهناك عديد من مقاييس التشتت أو التغير يمكن استخدامها وإن كان الأكثر شيوعاً هو المدى ، الإنحراف المتوسط ، الإنحراف المعياري ، نصف المدى الربيعي ، مدى المئينات .

الإنحراف المتوسط : Mean Deviation

الإنحراف المتوسط أو متوسط الإنحرافات لمجموعة n من الأرقام X_1, X_2, X_3 ... X_n يعرف بما يلي :

$$\text{الإنحراف المتوسط} = M.D = \frac{\sum_{i+1}^{n} |X_i - \bar{X}|}{n} = \frac{\Sigma\ |X - \bar{X}|}{n} \qquad (1)$$

حيث $\bar{X}$ هو الوسط الحسابي للأرقام و $| X_i - \bar{X} |$ هو القيمة المطلقة لانحرافات القيمة X_i عن $\bar{X}$. Absolute Deviation. القيمة المطلقة لرقم هو الرقم بدون الإشارة المرافقة له ويعبر عن ذلك بخطين رأسيين يوضعان حول الرقم وعلى ذلك فإن :

$$|-4| = 4,\ |+3| = 3,\ |6| = 6,\ |-0.84| = 0.84$$

مثال : أوجد متوسط الإنحرافات لمجموعة الأرقام 2, 3, 6, 8, 11

$$\text{الوسط الحسابي} = \bar{X} = \frac{2 + 3 + 6 + 8 + 11}{5} = 6$$

$$\text{الانحراف المتوسط} = M.D. = \frac{|2-6|+|3-6|+|6-6|+|8-6|+|11-6|}{5}$$

$$= \frac{|-4|+|-3|+|0|+|2|+|5|}{5}$$

$$= \frac{4+3+0+2+5}{}$$

$$= \frac{14}{5} = 2.8$$

إذا كانت $X_1, X_2, ..., X_k$ تحدث بتكرارات $f_1, f_2, ..., f_k$ على الترتيب فإن الانحراف المتوسط يمكن كتابته على صورة

$$\text{الانحراف المتوسط} = M.D. = \frac{\sum_{i=1}^{k} f_i |X_i - \bar{X}|}{n} = \frac{\sum f |X - \bar{X}|}{n} \quad (2)$$

حيث أن $n = \sum_{i=1}^{k} f_i = \sum f$ وهذه الصيغة مفيدة للبيانات المجمعة حيث X_i S تمثل مراكز الفئات و f_íS تمثل التكرارات المقابلة لهـا.

في بعض الأحيان يعرف الانحراف المتوسط بدلالة القيمة المطلقة للإنحرافات عن الوسيط أو غيره من المتوسطات بدلاً من الوسط الحسابي

خاصية هامة المجموع $\sum_{i=1}^{n} |X_i - a|$ يكون أقل ما يمكن عندما تكون a هي الوسيط ، بمعنى أن متوسط انحرافات القيم عن الوسيط يكون أقل ما يمكـــن .

لاحظ أنه قد يكون من الأنسب استخـدام التعبـير متوسـط القيم المطلقـة للإنحرافات MAD للتعبير عن الإنحراف المتوسط .

الإنحراف المعياري والتباين Standard Deviation and Variance

الانحراف المعياري Standard Deviation

الإنحراف المعياري لمجموعة من n رقم $X_1, X_2, ..., X_n$ يعبر عنها بالرمز S تعرف بما يلي

$$S = \sqrt{\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n-1}} = \sqrt{\frac{\Sigma x_i^2}{n-1}} \quad (3)$$

حيث أن x_i تمثل انحرافات كل رقم X_i عن $\bar{X}$.

وعلى هذا فإن S هي جذر متوسط مربعات انحرافات القيم عن وسطها الحسابي ويسمى أحياناً جذر متوسط مربع الإنحراف .

إذا كانت $X_1, X_2, X_3..., X_k$ تحدث بتكرارات $f_1, f_2, f_3, ..., f_k$ على الترتيب فإن الإنحراف المعياري يمكن كتابته على صورة

$$S = \sqrt{\sum_{i=1}^{k} \frac{f_i(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{\Sigma f(X - \bar{X})^2}{n-1}} = \sqrt{\frac{\Sigma fx^2}{n-1}} \quad (4)$$

حيث $n = \sum_{i=1}^{k} f_i = \Sigma f$ وهذه الصيغة مفيدة في حالة البيانات المجمعة

التبايـــن : Variance

تباين مجموعة من البيانات يعرف بأنه مربع الإنحراف المعياري وبهذا يعرف بـ S^2 في (3) و(4)

ويجب التمييز بين الإنحراف المعياري للمجتمع والإنحراف المعياري لعينة مسحوبة من هذا المجتمع ، لذا فإننا نستخدم دائماً الرمز S للأخير والرمز σ

للأول . وبهذا فإن S^2, σ^2 يمثلان تباين العينة وتباين المجتمع على الترتيب .*

طريقة مختصرة لحساب الإنحراف المعياري : -

المعادلات (3) و(4) يمكن كتابتها على الترتيب في الصيغ المتكافئة التالية :

$$(5) \qquad S = \sqrt{\frac{\sum_{i=1}^{n} X^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}}$$

$$(6) \qquad S = \sqrt{\frac{\sum_{i=1}^{k} f_i X^2 - \frac{\left(\sum_{i=1}^{k} f_i X_i\right)^2}{n}}{n-1}}$$

وإذا كانت $d_i = X_i - A$ هي انحرافات X_i عن ثابت اختياري A فالنتائج (5) و(6) تصبح على الترتيب.

$$(7) \qquad S = \sqrt{\frac{\sum_{i=1}^{n} d_i^2 - \frac{\left(\sum d_i\right)^2}{n}}{n-1}}$$

$$(8) \qquad S = \sqrt{\frac{\sum_{i=1}^{k} f_i d_i^2 \frac{\left(\sum_{i=1}^{k} f_i d_i\right)^2}{n}}{n-1}}$$

---

* ملحوظة : يعرف التباين للمجتمع والذي يرمز له بالحرف σ^2

كما يلي $\sigma^2 = \sum_{i=1}^{N} \frac{(X_i - \mu)^2}{N}$

أما إذا كانت مشاهدات المجتمع تكرر بتكرارات $f_1, f_2, \ldots, f_k$ فإن

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i (X_i - \mu)^2}{\sum f_i} = \frac{\sum f_i (X_i - \mu)^2}{N}$$

وتكون S^2 هي قيمة تقديرية للتبـاين σ^2 ومـن ثم فإن الإنحـراف المعياري للمجتمع هو σ وتقديره هو S .

وعندما تجمع البيانات في توزيع تكراري تتساوى فيه اطوال الفئات حيث يساوي طول الفئة C ، فإن

$$d_i = Cu_i \text{ أو } X_i = A + Cu_i$$

وعليه تصبح المعادلة (8) كما يأتي :

$$(9) \qquad S = C\sqrt{\frac{\sum_{i=1}^{k} f_i u_i^2 - \frac{(\sum_{i=1}^{k} f_i u)^2}{n}}{n-1}} = C\sqrt{\frac{\Sigma fu^2 - \frac{(\Sigma fu)^2}{n}}{n-1}}$$

والصيغة الأخيرة تعطى طريقة مختصرة جداً لحساب الإنحراف المعياري ويجب استخدامها للبيانات المجمعة إذا كانت أطوال الفئات متساوية وهذه تسمى بطريقة الترميز وهي مماثلة بالضبط للطريقة المستخدمة في حساب الوسط الحسابي من البيانات المجمعة في الفصل السابق .

خصائص الإنحراف المعياري :

1 ـ الإنحراف المعياري يمكن تعريفه كالآتي

$$(10) \qquad S = \sqrt{\sum_{i=1}^{n} \frac{(X_i - a)^2}{n-1}}$$

حيث a أي وسط بالإضافة للوسط الحسابي . ومن كل هذه الإنحرافات المعيارية ، نجد أن أصغرها يمكن الحصول عليه عندما نأخذ $a = \bar{x}$ وذلك حسب الخاصية (ب) من خصائص الوسط الحسابي في الفصل السابق .

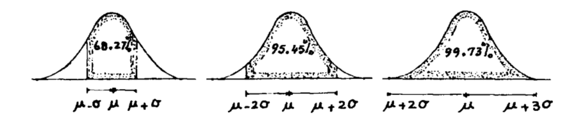
2 ـ في التوزيع الطبيعي (وهو أحد التوزيعات الإحصائية المهمة) نجد أن :

أ ـ 68.27% من الحالات تقع بين $\mu + \sigma$ و $\mu - \sigma$ (اي على بعد انحراف معياري واحد على كل جانب من الوسط الحسابي)

ب ـ 95.45% من الحالات تقع بين $\mu + 2\sigma$ و $\mu - 2\sigma$ (اي على بعد انحرافين معياريين على كل جانب من الوسط الحسابي).

جـ ـ 99.73% من الحالات تقع بين $\mu + 3\sigma$ و $\mu - 3\sigma$ (اي على بعد ثلاثة انحرافات معيارية على كل جانب من الوسط الحسابي).

(أنظر الأشكال الثلاثة أدناه)



3 ـ إذا افترضنا أن مجموعتين مكونتين من N_1 , N_2 رقم (أو توزيعـان تكراريان ومجموع تكرارتهما هما N_1, N_2) وتباينهما معطى بـ σ_1^2 , σ_2^2 على الترتيب ولهما نفس الوسط ، فإن التباين المشترك أو المجمع للمجموعتـين (أو للتوزيعـين التكراريين) هو :

$$\sigma^2 = \frac{N_1 \sigma_1^2 + N_2 \sigma_2^2}{N_1 + N_2} \qquad (11)$$

استعمالات الإنحراف المعيـاري

هنالك نظرية إحصائية تعرف بنظرية شبشف Chebychev تقول بأن نسبـة البيانات التي تقع بين الـوسط μ وبين k انحراف معياري تساوي على الأقل ($1 - \frac{1}{k^2}$) وعليه فإنه من المؤكد أن $\frac{3}{4}$ البيانات لأي مجموعة بيانات تقـع على بعـد

انحرافين معياريين 2σ لأن $3/4 = 1-1/4 = 1-\frac{1}{k^2}$ وأن $\frac{8}{9}$ من البيانات لأي مجموعة بيانات لا بد أن تقع على بعد ثلاثة انحرافات معيارية σ 3 لأن

$$1-\frac{1}{k^2} = 1-1/9 = 8/9$$

مثال أ: إذا كان االوسط الحسابي لمجموعتين من البيانات يبلغ 240 ($\mu=$) وأن الإنحراف المعياري للمجموعة الأولى قدره 50 $\sigma_1=$ أما الثاني فقدره 20 $\sigma_2=$ فإننا نستنتج من النظرية السابقة بأن % 75 من البيانات للمجموعة الأولى يقع بين 140 و 340 (أي على بعد انحرافين معيارين (50، 2) من الوسط) أما بالنسبة للمجموعة الثانية فإن 75% من البيانات تقع بين نفس الرقمين أي بين 200 و 280 وعليه فإن هذا المثال يوضح تأثير قيمة الإنحراف المعياري على تمركز البيانات حول الوسط الحسابي.

مثال ب: أجرت احدى شركات التأمين امتحانين للتدقيق في اختيار المتقدمين لها وقد كانت نتائج أحد المتقدمين كما يلي : أحرز 135 نقطة في الامتحان الأول و265 نقطة في الإمتحان الثاني . عليه فإنه يبدو للوهلة الأولى بأن هذا المتقدم قد أحرز نتيجة أفضل في المادة الثانية منها من الأولى . ولكن إذا عرفنا بأن الانحراف المعياري للمادة الأولى كان 15 $\sigma_1=$ وفي المادة الثانية كان 30 $\sigma_2=$ وأن الوسط الحسابي في المادة الأولى كان 100 $\mu_1=$ وفي المادة الثانية كان 250 $\mu_2=$ فإنه يتضح لنا أنه قد أحرز ($\frac{135-100}{15}$) = $2\frac{1}{3}$ قيمة معيارية في المادة الأولى.

بينما أحرز $\frac{265-250}{30}$ = $\frac{1}{2}$ قيمة معيارية في المادة الثانية وعليه فأن إنجازه في المادة الأولى كان أفضل منه في المادة الثانية.هذا المثال يوضح لنا أن ذكر البيانات دون معرفة المتوسط والإنحراف المعياري لا يؤدي الى

تحليل دقيق للبيانات وعليه فإننا غالباً ما نعبر عن البيانات التـي لدينـا بوحدات الإنحراف المعياري بمعنى آخر إذا كان لدينا بيان X فإننا نحول هذا البيان الى قيم معيارية يرمز لها بالحرف Z وقيمتها كما يلي

$$Z = \frac{X - \mu}{\sigma} \text{ or } Z = \frac{X - \bar{X}}{S}$$

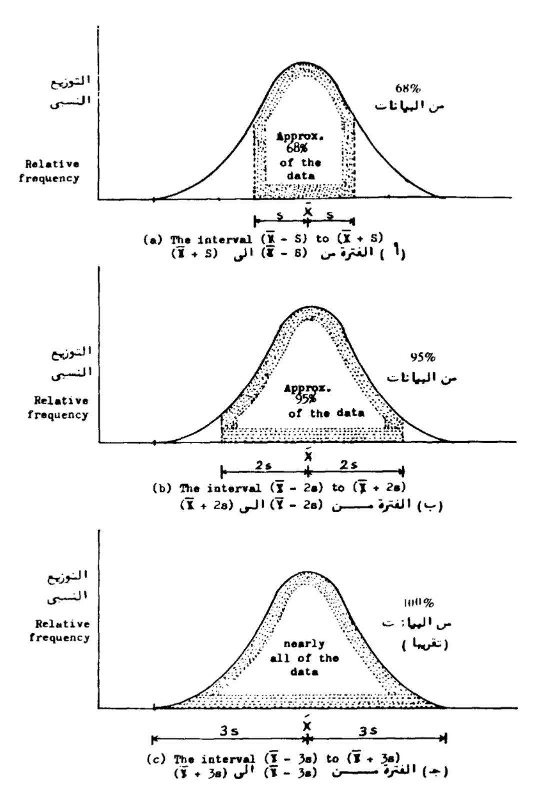
القانون التجريـــبي (الامبريـــقي) Empirical Rule

أثبتت الدراسات التجريبية (الامبريقيـــة) بأنه إذا كان لدينا توزيعاً متماثلاً (أو ما يقاربـه) : -

أ) فإن الفترة من $(\bar{X} - S)$ الى $(\bar{X} + S)$ تحتوي على ما يقارب 68% من البيانات .

ب) أن الفترة من $(\bar{X} - 2S)$ الى $(\bar{X} + 2S)$ تحتوي على ما يقارب 95% من البيانات .

جـ) وأن الفترة من $(\bar{X} - 3S)$ الى $(\bar{x} + 3S)$ تحتوي على ما يقارب 100% من البيانات أي أن كل البيانات تقع تقريباً ما بين $(\bar{X} - 3S)$ و $(\bar{X} + 3S)$



Graphical Representation of the Empirical Rule.

التمثيل البياني للقانون التجريبي

مثـــال:

الجدول يوضح أوزان 100 طالب في جامعة ما:

عدد الطلبـة	الأوزان (كيلوجرامات)
5	60 - 62
18	63 - 65
42	66 - 68
27	69 - 71
8	72 - 74
100	

المطلوب إيجاد الإنحراف المعياري

(أ) بالطريقة المطولة.

(ب) بطريقة الترميز (أي الطريقة المختصرة).

ومن ثم أوجد التباين.

الحـــل :

$$S = \sqrt{\frac{\Sigma f(X - \overline{X})^2}{n - 1}}$$

(أ) لاستعمال القانون يمكن ترتيب الجدول الآتي : -

$f_i(X_i - X)^2$	f_i	$(X_i - \overline{X})^2$	$X_i - \overline{X} =$ $X - 67.45$	X_i	
208.0125	5	41.6025	− 6.45	61	60 - 62
214.2450	18	11.9025	− 3.45	64	63 - 65
8.5050	42	0.2025	− 0.45	67	66 - 68
175.5675	27	6.5025	2.55	70	69 - 71
246.4200	8	30.8025	5.55	73	72 - 74
$\Sigma f(X_i - \overline{X})^2$ = 852.7500	$n = \Sigma f_i =$ 100				

$$\overline{X} = \frac{\sum_{i=1} fX}{n} = \frac{\Sigma fX}{n} = \frac{6745}{100} = 67.45$$

$$S = \sqrt{\frac{\Sigma f\,(X - X)^2}{n - 1}} = \sqrt{\frac{852\ .7500}{100 - 1}} = \sqrt{8.6136} = 2.93$$

$$S^2 = 8.6$$

(ب) الطريقـــة المختصـــرة:

$f_iX_i^2$	f_i	X_i^2	X_i	الوزن Kg
18605	5	3721	61	60 - 62
73728	18	4096	64	63 - 65
188538	42	4489	67	66 - 68
132300	27	4900	70	69 - 71
42632	8	5329	73	72 - 74
$\Sigma fX^2 =$ 455803	$n = \Sigma f =$ 100			

وبتطبيق الصيغة المختصرة

$$S = \sqrt{\frac{\Sigma fX^2 - \frac{(\Sigma fX)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{455803 - 100\ (67.45)^2}{100-1}} = \sqrt{8.6136} = 2.93kg$$

$$S^2 = (2.93)^2 = 8.6$$

المدى ونصف المدى : Range and Midrange

مدى مجموعة من الأرقام هو الفرق بين أكبر رقم وأقل رقم من المجموعة.

مثال: مدى المجموعة 12 ,10 ,8 ,5 ,5 ,3 ,3 ,2 هو 10 = 2 − 12

في بعض الأحيان يعطى المدى بذكر أقل وأكبر رقم. في المثال السابق على سبيل المثال يمكن تحديد المدى من 2 الى 12 أو 12-2 أما نصف المدى هو عبارة عن متوسط المدى أي الفرق بين أكبر رقم وأقل رقم مقسوماً على 2 .

ونصف المدى في هذا المثال يساوي 5 = 2/ 10

مقاييس أخرى للتشتت :

نصف المدى الربيعي أو الإنحراف الربيعي: Interquartile Range لمجموعة من البيانات هو :

$$\frac{Q_3 - Q_1}{2}$$

حيث أن Q_1 هو الربيع الأول و Q_3 هو الربيع الثالث للبيانات ويستخدم

المدى الربيعي $Q_3 - Q_1$ في بعض الأحيان بدلاً من نصف المدى الربيعـي كمقياس شائع للتشتت.

ألتشتت المطلق أو النسبي ومعامل الإختلاف Coefficient of Variation أو C.V.

التغير الفعلي أو التشتت كما نحصل عليه من الإنحراف المعيـاري أو غيره من مقـاييس التشتت يسمى بالتشتت المطلق . ولكن تغير أو تشتت 1 متر عن قياس مسافة 1000 متر يختلف في تأثيره عن نفس تغير 1 متر في مسافة 20 متر . ومقياس هذا التأثير نحصل عليه بالتشتت النسبي ويعرف بما يلي :

$$\text{التشتت النسبي} = \frac{\text{التشتت المطلق}}{\text{المتوسط}}$$

إذا كان التشتت المطلق هو الإنحراف المعياري والمتوسط هو $\bar{X}$ فإن التشتت النسبي يسمى بمعامل الإختلاف أو معامل التشتت ويعرف كالآتي :

$$C.V = \frac{S}{\bar{\bar{X}}} \times 100 \qquad (13)$$

المتغير المعياري والدرجات المعيارية : The Standardized and the standard Unit

المتغيـر

$$Z = \frac{X - \bar{X}}{S} \qquad (14)$$

والـذي يقيس الانحراف عن الـوسط الحسابي بـوحدات من الانحـراف المعياري ، يسمى وحدة معيارية أو درجة معيارية . وهذه لهـا دلالة كبيـرة عند المقارنة بين التوزيعات (Distributions) الاحصائية

طريقة سريعة لتقدير الإنحراف المعياري:

إذا أردنا الحصول على قيمة تقديرية للإنحراف المعياري فإننا غالباً ما نأخذ القيمة $\frac{\text{المدى}}{4} = \frac{Range}{4} = S$

مثـال :

أجرى طلبة إحدى كليات التجارة دراسة لمعرفة الزيادة في أسعار احدى السلع الإستهلاكية ولإجراء هذه الدراسة فقد قاموا بالحصول على بيانات بهذه الزيادات لأربعة وعشرين عينة مختلفة في مدة ستة أشهر وقد كانت الزيادات كما موضحة في الجدول (1) أدناه . هذا وكانت الدراسات السابقة قد دلت على أن الزيادات في الأسعار لهذه السلعة تأخذ شكلاً متماثلاً . لذا فقد كان أحد أهداف الدراسة بحث فيما لو كان شكل توزيع الزيادات قد تغيّر أم لا .

الإجابـة :

يتضح من الجدول (1) أن الوسط الحسابي للزيادة هو $\bar{X} = 118.33$ وأن الإنحراف المعياري هو $s = 15.08$ وللإجابة على السؤال حول الشكل لتوزيع الزيادات فإننا يمكن أن نستعين بالقانون التجريبي والذي يقول أن 68% (تقريبا) من البيانات يقع بين $(\bar{X} - s)$ و $(\bar{X} + s)$ بالتعويض لقيم المتوسط الحسابي والإنحراف المعياري يتضح لنا أن هذه الفترة تقع ما بين $(118.33 - 15.08)$ و $(15.08 + 118.33)$أي بين 103.25، 133.41 وبالرجوع الى الجدول رقم (2) يتضح لنا أن ثمانية عشر من الأربعة وعشرين زيادة تقع في هذه الفترة . أي أن 75% من الزيادات تقع في هذه الفترة . وباستعمال نفس القانون فإنه من المفروض أن تقع 95% (تقريبـا) من الزيادات ما بين $(\bar{X} - 2s), (\bar{X} + 2s)$

جدول (1) الزيادة في الأسعار

100	121	130	129
150	116	120	117
154	125	110	119
130	115	125	123
90	109	100	120
92	112	115	118

جدول (2) الزيادة في الأسعار بعد وضعها في مصفوفة تصاعديـة

90	112	119	125
92	115	120	129
100	115	120	130
100	116	121	130
109	116	123	150
110	118	125	154

$\bar{X}$ 118.33 و S = 15.08

أي ما بـين ((118.33 − 2(15.08) و ((118.33 + 2(15.08) أي ما بـين 88.17 و 148.49 . وبالرجوع الى البيانات في جدول رقم (2) يتضح لنا أن 22 من الـ24 زيادة تقع في هذه الفترة . أي أن 92% من الزيادات تقع في هذه الفترة.

وكذلك الحال فإن القانون يفترض أن تقع 99.9% (تقريبا) من الزيادات ما بين (X− 3s) و (X+3s) وبالتعويض أيضا نحصل على (118.33− 3(15.08)) و (3(15.08)+118.33) أي أن الفترة هي 73.09 إلى 163.57 وبالرجوع إلى البيانات في الجدول رقم (2) يتضح لنا أن كل البيانات أي 100% تقع في هذه الفترة .

الخلاصـة :

إن هذه النتائج توضح لنا أن الزيادة في أسعار هذه السلع ما زالت تأخذ نفس الشكل المتماثل كما كان في السابـق .

تنبيـــه :

أن هذه النتيجة لا تقول بأن الزيادات في هذه المرة كانت نفس الزيادات في السابق بمعنى أنها لم تقل أن متوسط الزيادة هو المتوسط في السابـق وأن انحرافهـا المعياري هو نفس الانحراف المعياري بل أن كل ما توضحه هو أن شكل توزيـع الزيادات مازال كما كان في السابق .

تمارين :

(1) البيانات التالية أخذت من ملفات شركة تأمين على الحياة ، سجلت فيها عدد الأسابيع المتعاقبة والتي دفعت فيها الشركة تعويضات لـ ١٤ من ضحايا الذبحة الصدريـة :

7	16	8	23	14	21	9
20	10	32	10	13	27	20

المطلوب حساب الآتي :

(أ) متوسط عدد الأسابيع التي دفعت فيها التعويضات .

(ب) أوجد الوسيط لعدد الأسابيع التي دفعت فيها التعويضات .

(جـ) أوجد المنوال .

(د) أوجد الإنحراف المعياري لعدد الأسابيع التي دفعت فيها التعويضات .

(هـ) أحسب متوسط القيمة المطلقة للإنحراف MAD لعـدد الأسـابيع التـي دفعت فيها التعويضـات .

(2) عندما تتعطل ماكينة لفترة من الزمن في مصنع ما نتيجة لعطب في أثناء ساعات العمل الرسمية فإن هذا الزمن يسمى «بالوقت الضائع»·التوزيع التالي لعينـة تتكون من 100 من الأوقات الضائعة لماكينة معينة (لأقرب دقيقة) .

الوقت الضائع (دقيقة)	التكرارات (عدد مرات التعطل)
0 — 9	3
10 — 19	13
20 — 29	30
30 — 39	25
40 — 49	14
50 — 59	8
60 — 69	4
70 — 79	2
80 — 89	1

المطلوب حساب الآتي : -

(أ) متوسط الأوقات الضائعة في هذا التوزيع .

(ب) الانحراف المعياري للأوقات الضائعة في هذا التوزيع .

(جـ) الوسيـط.

(د) معامل التشتت .

(هـ) الربيع الأول والثالث.

(و) قدر الإنحراف المعياري بطريقة تختلف عن (ب).

(3) كانت نتائج أحد الامتحانات للشهادةالثانوية في إحدى الدول كما هي موضحة في الجدول أدناه.

المادة	المتوسط	الانحراف المعياري
الرياضيات	67	12
الفيزيــاء	75	8
الكيميــاء	70	18
الجغرافيــا	65	15
العربي	72	16
الانجليــزي	53	20
كل المواد	67	17

أولا: احسب التشتت النسبي لهذه المواد كل على حدة.

ثانياً: أ ـ أحرز أحد الطلبة النتيجة الآتية:

الرياضيات 80، الفيزياء 85، الكيمياء 88، الجغرافيا 82، العربي 88 وفي الانجليزي 74.

(1) في أي المواد كان أداؤه أفضل.

(2) كم كان متوسط الدرجــات.

(3) إذا علمت بأنه أعاد هذا الامتحان للمرةالثانية واحرز 85 درجة في الـرياضيات حيث كان متوسط الدرجـات فيهـا 70 وانحرافهـا المعياري 16. في أي الإمتحانين كان درجـة أداؤه أفضـل في هذه المادة.

الفصـــل الرابع

الارتباط والانحدار

CORRELATION AND REGRESSION

الفصـــل الرابع

الارتباط والانحدار

CORRELATION AND REGRESSION

وصف العلاقة بين متغيريـــن :

يختص الإنحدار البسيط بدراسة العلاقة بين متغيرين على هيئة علاقة دالية بحيث يمكن التنبؤ منها عن أحد المتغيرين بمعلومية المتغير الآخر . فإذا كانت هناك علاقة انحدار بسيط بين متغيرين فإن أحد المتغيرين يعرف بالمتغير المستقل Independent Variable ويرمز له بالرمز (X) بحيث أن أي تغير في المتغير المستقل يتسبب في تغير المتغير الآخر والذي يعرف بالمتغير التابع dependent Variable ويرمز له بالرمز (Y).

أما إذا كانت العلاقة بين المتغيرين ليست علاقة دالية بل علاقة ترابط فقط أي أن التغير في أحدهما لا يسبب التغير في الآخر فإن مهمة الإحصاء هي قياس درجة الترابط بين المتغيرين . وتعرف العلاقة بين المتغيرين في هذه الحالة بالإرتباط Correlation ولكن في حالة العلاقة الدالية فإن علاقة الارتباط دائماً موجودة.

ومن أمثلة الارتباط فقط ، العلاقة بين الأخ والأخت فإن أي تغير في أحدهما لا يسبب تغير في الآخر ولكن المتغيرين مرتبطين عن طريق الاباء، كذلك العلاقة بين

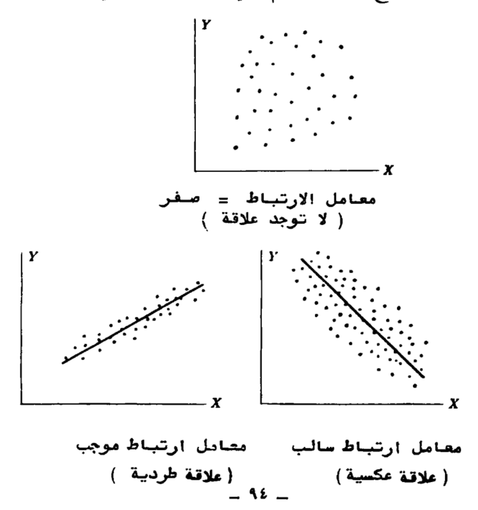
طول وعرض أوراق النبات . فإن أي تغير في إحداهما لا يسبب تغيراً في الأخرى ولكن الاثنين مرتبطان بحجم النبات .

ويرمز للمتغيرين في حالة الإرتباط بالرمز X_1, X_2 أو X و Y دون أن يكون الأول المتغير المستقل والثاني المتغير التابع .

عموماً إذا كان عدد المتغيرات اثنين فقط فإننا نتحدث عن الإرتباط البسيط والانحدار البسيط أمّا إذا كان هناك أكثر من متغيرين فإننا نتحدث عن الارتباط المتعدد والانحدار المتعدد ،وفي هذا الفصل سوف نتناول الإرتباط البسيط فقط .

مفهوم الإرتباط ـ التحليل البياني:

يقاس الارتباط بين X_1 و X_2 أو X و Y بعمل رسم انتشار Scatter Diagram وذلك بتوقيع النقط على رسم بياني ويمثل المحور الأفقي X_1 أو X والمحور



معامل الارتباط = صفر
(لا توجد علاقة)

معامل ارتباط سالب
(علاقة عكسية)

معامل ارتباط موجب
(علاقة طردية)

الرأسي X_2 أو Y . فإذا انحصرت معظم النقط داخل قطاع ضيق دلّ ذلك على ارتباط قوي أما إذا ما انحصرت النقط داخل دائرة دل ذلك على ضعف الإرتباط.

والشكل أعلاه يوضح ثلاث حالات من الارتباط ورسم الإنتشار لكل حالـة. عيب هذه الطريقة أنها تقريبية ولا تصلح للدراسات الدقيقة .

معادلة الإنحدار البسيط باستخدام المربعات الصغرى :

سنتعرض أولا لمدى جودة تعبير خط مستقيم عن العلاقة بين متغيرين لهذا فإننا نحتاج اولا لمعادلات الانحدار باستخدام المربعات الصغرى لخط انحدار Y على X .

$$(1) \qquad Y = a_0 + a_1X + \varepsilon$$

حيث نحصل على قيم a_0 , a_1 من المعادلات الاعتيادية .

$$(2) \qquad \Sigma Y = a_0N + a_1 \Sigma X$$

$$(3) \qquad \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2$$

ومنها :

$$(4) \qquad a_0 = \frac{(\Sigma Y)(\Sigma X^2) - (\Sigma X)(\Sigma XY)}{N \Sigma X^2 - (\Sigma X)^2} = \mu_y - a_1 \mu_x$$

$$(5) \qquad a_1 = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N \Sigma X^2 - (\Sigma X)^2} = \frac{\sigma_{xy}}{\sigma_{x^2}}$$

كذلك فإن خط انحدار X على Y هو

$$(6) \qquad X = b_0 + b_1Y + \eta$$

حيث نحصل على قيم b_1, b_0 من المعادلات الاعتيادية

$$\Sigma X = b_0 N + b_1 \Sigma Y \tag{7}$$

$$\Sigma XY = b_0 \Sigma Y + b \Sigma Y^2 \tag{8}$$

ومنهـــا :

$$b_0 = \frac{(\Sigma X)(\Sigma Y^2) - (\Sigma Y)(\Sigma XY)}{N \Sigma Y^2 - (\Sigma Y)^2} = \mu_x - b_1 \mu_y \tag{9}$$

$$b_1 = \frac{N \Sigma XY - (\Sigma X)(\Sigma Y)}{N \Sigma Y^2 - (\Sigma Y)^2} = \frac{\sigma_{xy}}{\sigma_{y^2}} \tag{10}$$

خط الإنحـــدار

ايجاد تقديرات ثوابت خط الإنحـــدار باستعمال المصفوفـات

إن خط الإنحدار للعينـــة هـو

$$Y_i = \alpha + \beta X_i + \varepsilon_i , \quad i = 1, 2, ..., n$$

وعليه يمكن كتابة هذا في صيغة المصفوفة التالية

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \alpha + & \beta X_1 \\ \alpha + & \beta X_2 \\ \alpha + & \beta X_3 \\ & \vdots \\ \alpha + & \beta X_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \\ \varepsilon_n \end{bmatrix}$$

أي أن

$$\underset{n\times 1}{\begin{bmatrix} Y_1 \\ Y \\ Y_3 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix}} = \underset{n\times 2}{\begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \\ \vdots & \\ 1 & X_n \end{bmatrix}} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}_{2\times n} + \underset{n\times 1}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}}$$

$$\underline{Y} = X\,\underline{\beta} + \underline{\varepsilon}$$

حيث أن

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \underline{\beta} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

وعليه فإنه يمكن تقدير قيمة

$$\hat{\underline{\beta}} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{ب}\ \underline{\beta} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

أي أن خط الإنحدار المقدر هو

$$\underline{Y} = X\hat{\underline{\beta}}$$

ومن ثم فان

$$X'\underline{Y} = X'X\,\hat{\beta}$$

$$(X'X)^{-1}X'\underline{Y} = (X'X)^{-1}X'X\underline{\hat{B}}$$

أي أن

$$\underline{\hat{\beta}} = (X'X)^{-1}X'\underline{Y}$$

حيث أن :

$$X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \underline{\hat{\beta}} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_1 & X_2 & X_3 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} Y_1 + Y_2 + \dots + Y_n \\ X_1Y_1 + X_2Y_2 + X_3Y_3 + \dots + X_nY_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{bmatrix}$$

$$[X'X]=\begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$$

$$=\begin{bmatrix} 1 \times 1 + 1 \times 1 + \dots + 1 \times 1, X_1 + X_2 + \dots + X_n \\ \\ X_1 + X_2 + \dots X_n, X_1{}^2 + X_2{}^2 + \dots + X_n^2 \end{bmatrix}$$

$$X'X = \begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X^2_i \end{bmatrix}$$

$$|X'X| = n \sum X^2_i - (\sum X_i)^2$$

$$\therefore (X'X)^{-1} = \frac{1}{|X'X|} \begin{bmatrix} \sum_{i=1}^{n} X^2_i & -\sum X_i \\ \\ -X_i & n \end{bmatrix}$$

$$= \frac{1}{n \sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X^2_i & -\sum X_i \\ \\ -\sum X_i & n \end{bmatrix}$$

$$= \frac{1}{n(n-1)S_x^2} \begin{bmatrix} \sum X^2_i & -\sum X_i \\ \\ -\sum X_i & n \end{bmatrix}$$

$$\therefore (X'X)^{-1} X'Y = \frac{1}{(n-1)n\,S_x^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix} \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}$$

$$= \frac{}{(n-1)n\,S_x^2} \begin{bmatrix} (\Sigma X_i)^2 (\Sigma Y_i) - (\Sigma X_i)(\Sigma X_i Y_i) \\ -(\Sigma X_i)(\Sigma Y_i) + n\,\Sigma X_i Y_i \end{bmatrix}$$

$$\therefore \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \frac{1}{(n-1)n\,S_x^2} \begin{bmatrix} (\Sigma X_i)^2 (\Sigma Y_i) - (\Sigma X_i)(\Sigma X_i Y_i) \\ -(\Sigma X_i)(\Sigma Y_i) + n\,\Sigma X_i Y_i \end{bmatrix}$$

$$b = \hat{\beta} = \frac{n\Sigma X_i Y_i - (\Sigma X_i)(\Sigma Y_i)}{(n-1)\,n\,S_x^2}$$

$$= \frac{n\left[\Sigma X_i Y_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}\right]}{(n-1)\,n\,S_x^2}$$

$$b = \hat{\beta} = \frac{n\,S_{XY}}{n\,S_x^2} = \frac{S_{xy}}{S_x^2}\ ; \qquad \boxed{b = \hat{\beta} = \frac{S_{xy}}{Sx^2}}$$

$$a = \frac{(\Sigma X_i^2)\,\Sigma Y_i - (\Sigma X_i)\,\Sigma X_i Y_i}{(n-1)\,n\,S_x^2}$$

$$= \left[(\Sigma X_i^2)(\Sigma Y_i) - \frac{(\Sigma X_i)^2}{n}(\Sigma Y_i) + \frac{(\Sigma X_i)^2}{n}(\Sigma Y_i) - (\Sigma X_i)\,\Sigma X_i Y_i \right] / \, n^2 S_x^2$$

$$= \frac{1}{(n-1)nS_x^2} \left[(\Sigma Y)\left(\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right) - (\Sigma X_i)\left((\Sigma X_i Y_i) - (\Sigma X)(\Sigma Y)/n\right)\right]$$

$$a = \frac{\Sigma Y_i}{(n-1)\,n\,S_x^2}\,[n-1]\,S_x^2 - \frac{\Sigma X_i}{(n-1)nX^2}\,[n-1]\,S_{XY}$$

$$a = \overline{Y} - b\,\overline{X}$$

$$\boxed{a = \hat{\alpha} = \overline{Y} - b\overline{X}}$$

أي أن

$$\boxed{\begin{array}{l} a = \hat{\alpha} = \overline{y} - b\overline{X} \\ b = \hat{\beta} = \dfrac{Sxy}{S^2x} \end{array}} \qquad (11)$$

$$\hat{\beta} = \begin{bmatrix} \hat{\hat{\alpha}} \\ \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} a \\ \\ b \end{bmatrix} \quad \begin{bmatrix} \overline{Y} - b\overline{X} \\ \\ \dfrac{Sxy}{S^2x} \end{bmatrix}$$

وهي نفس القيم التي حصلنا عليها باستعمال نظرية المربعات الصغرى.
إذا ما أخذنـا $S_y, S_x, S_{xy}, \overline{Y}, \overline{X}$ كقيم تقديريـــة للمعالم $\sigma_y, \sigma_x, \sigma_{xy}, \mu_y, \mu_x$ على التــوالي.

المعادلات (1) و (6) يمكن كتابتهما أيضاً على الصورة التالية:

$$y = \left(\frac{\Sigma xy}{\Sigma x^2}\right) x, \quad x = \left(\frac{\Sigma xy}{\Sigma y^2}\right) y \qquad (12)$$

أي أن $y = a_1 x, \quad x = b_1 y$

حيث $y = Y - \overline{Y} \quad , \quad x = X - \overline{X}$

وتتساوى معادلتا الانحدار في حالة وحيدة فقط إذا كانت جميع النقاط في شكل الانتشار تقع على خط ذو ميل 45، درجة وفي هذه الحالة فإن هناك إرتباطا خطيا بين X و Y.

الخطأ المعياري للتقديرات : Standard Error of Estimate

إذا كانت $\hat{Y}$ تمثل تقديراً لقيمة Y المقابلة لقيمة معينة لـ X مستخدمين المعادلة (١)، فإن مقياس الانتشار حول خط انحدار Y على X نحصل عليه من الكمية.

$$S_{y.x} = \sqrt{\frac{(Y-\hat{Y})^2}{n-2}} \qquad (13)$$

حيث $S_{y.x}$ هي قيمة تقديرية لـ $\sigma_{y.x}$ (الإنحراف المعياري للمشاهدات Y_i بعد أخذ قيم X_i بالإعتبار) وتسمى بالخطأ المعياري لتقدير Y على X

وإذا استخدمنا خط الإنحدار (6) فإن الخطأ المعياري لتقدير X على Y يعرف كالآتي:

$$S_{x.y} = \sqrt{\frac{(X-\hat{X})^2}{n-2}} \qquad (14)$$

حيث $S_{x.y}$ هي قيمة تقديرية لـ $\sigma_{x.y}$ (الانحراف المعياري للمشاهدات X_i بعد أخذ قيم Y_i بالإعتبار) وبشكل عام فإن $S_{x.y}$ يساوي $S_{y.x}$

المعادلة (13) يمكن كتابتها على الصورة:

$$S^2_{y.x} = \frac{\Sigma Y^2 - \hat{a}_0 \Sigma Y - \hat{a}_1 \Sigma XY}{n-2} \qquad (15)$$

والتي قد تكون أكثر ملائمة للحساب ويمكن الحصول على تعبير مماثل للمعادلة (14).

الإختلاف المفسر والإختلاف الغير مفسر:

Explained and unexplained variation

يعرف الإختلاف الكلي لـ Y بأنه $\sum_{i=1}^{n}(Y_i-\bar{Y}_i)^2$ أي مجموع مربعات انحرافات

قيم Y عن الوسط $\overline{Y}$ ويمكن كتابته على الصورة الآتيـة :

$$\Sigma (Y_i - \hat{Y}_i)^2 = \Sigma (Y - \hat{Y})^2 + \Sigma (\hat{Y} - \overline{Y})^2 \qquad (16)$$

ويسمى الحد الاول بالإختلاف غير المفسر والحد الثاني بالإختـلاف المفسر ، رهذه التسمية راجعة إلى أن الاختلافات $\hat{Y}_i - \overline{Y}$ لها نموذج محدد ، بينما الإختلافات $Y_i - Y_i$ تسلك سلوكا عشوائيا أو بصورة لايمكن التنبؤ بها .

معامل الإرتباط : Correlation Coefficient

النسبة بين الاختلافات المفسرة والاختلاف الـكلي تسمـى معامـل التحـديد Coefficient of Determination فإذا كانت الاختلافات المفسرة تساوي صفراً . أي أن الاختلاف الكلي جميعه غير مفسر، فإن هذه النسبة تساوي صفرا . أما إذا كانت الاختلافات الغيرمفسرة تساوي صفراً ، أي أن الاختلاف الـكلي جميعـه مفسرٌ فإن النسبـة تسـاوي واحـداً . وفي الحـالات الأخـرى تقـع هذه النسبـة بـين الصفـر والواحـــد .

بما أن النسبة دائماً غير سالبة ، فنرمز لها بالرمز r^2 حيث r هو معامل الارتباط ويعرف كالآتي :

$$r = \sqrt{\frac{\text{الاختلاف المفسر}}{\text{الاختلاف الكلي}}} = \pm\sqrt{\frac{\Sigma (\hat{Y} - \overline{Y})^2}{\Sigma (Y - \overline{Y})^2}} \qquad (17)$$

ويتراوح بين -1 و $+1$. العلامات ± تستخدم للارتبـاط الخطـي الموجـب والارتباط الخطي السالب . لاحظ أن r لا تمييزلها أي أنها لا تعتمد على الوحـدات المستخدمة .

باستخدام المعادلات (13)،(16) مع ملاحظة أن الإنحراف المعياري لـ Y هو :

$$S_y = \sqrt{\frac{\Sigma (Y - \overline{Y})^2}{n-1}} \qquad (18)$$

نجد أن المعادلة (17) يمكن كتابتها بإهمال الاشارة كالآتي:

$$r=\sqrt{1-\frac{S^2_{y.x}}{S^2_y}} \quad \text{أو} \quad S_{y.x}=\sqrt{1-r^2}\,.\,S_y \qquad (19)$$

ويمكن إيجاد تعبيرات مماثلة اذا أبدلنا X و Y في حالة الإرتباط الخطي فإن تظل كما هي بصرف النظر عما اذا اعتبرنا X أو Y هو المتغير المستقل ، بهذا فإن r يعد مقياساً جيداً للإرتباط الخطي .

ملاحظة خاصة بمعامل الإرتباط:

التعاريف (17) أو (19) لمعامل الإرتباط تعاريف عامة ويمكن استخدامها للعلاقة الغير خطية وكذلك للعلاقة الخطية ، والإختلاف الوحيد هو أن $\hat{Y}$ تحسب من معادلة انحدار غير خطية بدلاً من معادلة الإنحدار الخطية وتحذف الاشارة السالبة وفي هذه الحالة تعتبر المعادلة (13) التي تعرف الخطأ المعياري للتقدير ، تعريفاً عاماً .

المعادلة (15) والتي تطبق في حالة الإنحدار الخطي فقط ، يجب تعديلها . فإذكانت المعادلة المقدرة ، على سبيل المثال هي:

$$Y = a_0 + a_1X_1 + a_2X_2 + ... + a_{k-1}X_{k-1} + \varepsilon \qquad (20)$$

فإن المعادلة (15) تستبدل بالمعادلة :

$$S^2_{\ y.x}= \frac{\Sigma Y^2 - \hat{a}_0\Sigma Y - \hat{a}_1 \Sigma X_1Y - ... - \hat{a}_{k-1}\Sigma X_{k-1}Y}{n-k} \qquad (21)$$

يجب التأكيد على أن قيمة r المحسوبة في أية حالة تقيس درجة العلاقة بالنسبة الى نوع المعادلة المفترضة . فإذا افترضنا معادلة خطية وإذا نتج عن المعادلة (17) أو (19) قيمة ل r تقترب من الصفر، فهذا يعني أنه لا يوجد تقريباً علاقة خطية بين

المتغيرات . ولكن هذا لا يعني أنه لا يوجد علاقة بين المتغيرات على الإطلاق ، حيث أنه قد يكون هناك بالفعل علاقة كبيرة غير خطية بين المتغيرات . وبصورة أخرى فإن معامل الإرتباط يقيس مدى جودة توفيق المعادلة المفترضة للبيانات . ما لم يوضح خلاف ذلك ، فإن مصطلح معامل الإرتباط يستخدم ليعني الإرتباط الخطي .

ويجب إيضاح أن وجود معامل ارتباط مرتفع (أي يقترب من 1 أو 1 -) لا يعني وجود علاقة تبعية مباشرة بين المتغيرات . فقد يكون هناك معامل ارتباط مرتفع بين عدد الكتب المنشورة في كل سنة وعدد مباريات الكرة الملعوبة في كل سنة مثل هذه الأمثلة يشار إليها بأنها ارتباط لا معنى له أو ارتباط زائف Spurious Correlation

صيغة عزم حاصل الضرب لمعامل الإرتباط الخطي : Product-Moment Formula

إذا افترضنا وجود علاقة خطية بين متغيرين ، فإن المعادلة (17) تصبح

$$r = \frac{\sum xy / n-1}{\sqrt{\frac{\sum x^2}{n-1} \cdot \frac{\sum y^2}{n-1}}} = \frac{S_{xy}}{\sqrt{S_x^2 \cdot S_y^2}} = \frac{S_{xy}}{S_x \, S_y} \qquad (22)$$

حيث $x = X - \bar{X}$, $y = Y - \bar{Y}$ هذه الصيغة والتي تعطى تلقائياً الإشارة المناسبة لـ r ، تسمى صيغة عزم حاصل الضرب وتظهر بشكل واضح التماثل بين Y, X.

لاحظ أن

$$S_{xy} = \frac{\sum xy}{n-1} \, , \, S_x = \sqrt{\frac{\sum x^2}{n-1}} \, S_y = \sqrt{\frac{\sum y^2}{n-1}} \qquad (23)$$

حيث S_y , S_x تعبر عن الإنحرافات المعيارية للمتغيرات Y, X على الترتيب، بينما S^2_x , S^2_y تعبر عن تبايناتهما والمقدار الجديد S_{xy} يسمى تغاير Y, X وباستخدام رموز

المعادلتين (22) ، (23) يمكن أن نكتب :

$$r = \frac{S_{xy}}{S_x S_y} \qquad (24)$$

لاحظ أن r لا تعتمد على وحدات قياس Y, X كذلك لا تعتمد على اختيار نقطة الأصل وهذه القيمة تسمى معامل ارتباط بيرســون . Pearson Correlation Coefficient

صيغة مختصرة للعمليات الحسابيــة :

الصيغة (22) يمكن كتابتها بصورة أخرى كالآتي :

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}} \qquad (25)$$

$$= \frac{\Sigma XY - (\Sigma X)(\Sigma Y)/n}{\sqrt{[\Sigma X^2 - (\Sigma X)^2/n][\Sigma Y^2 - (\Sigma Y)^2/n]}}$$

وهذه الصيغة تستخدم غالباً عند حساب r .

وبالنسبة للبيانات المجمعة في جدول لمتغيرين أو التوزيع التكراري لمتغيرين فإنه من الملائم استخدام الطريقة المختصرة وفي مثل هذه الحالة نجد أن المعادلة (25) يمكن كتابتها كالآتي :

$$r = \frac{n\Sigma fu_x u_y - (\Sigma f_x u_x)(\Sigma f_y u_y)}{\sqrt{(n\Sigma fu_x^2 - (\Sigma f_x u_x)^2][n\Sigma f_y u_y^2 - (\Sigma f_y u_y)^2]}} \qquad (26)$$

حيث f, f_x, f_y, u_x, u_y هي نفس الرموز في الفصل الثاني

لتسـهيل العمليات الحسـابية باستخـدام هذه الصيغـة ، نستخـدم جدول

ارتباط . أما بالنسبة للبيانات المجمعة ، فيمكن كتابة الصيغة (23) كالآتي :

$$S_{xy} = C_x C_y \left[\frac{\sum f u_x u_y - \frac{(\sum f_x u_x)(\sum f_y u_y)}{n}}{n-1} \right] \quad (27)$$

$$S_x = C_x \sqrt{\frac{\sum f_x u_x^2 - \frac{(\sum f_x u_x)^2}{n}}{n-1}} \quad (28)$$

$$S_y = C_y \sqrt{\frac{\sum f_y u_y^2 - \frac{(\sum f_y u_y)^2}{n}}{n-1}} \quad (29)$$

حيث C_x , C_y هو طول الفئة (مفترضاً أنها ثابتة) المقابلة للمتغيرات X, Y على الترتيـــب .

الصيغة (24)، يمكن إثبات أنها مكافئة للصيغة (26) إذا استخدمنا النتائـج (27) - (29) .

مثـال تطبيـقي : الارتباط Correlation

أجرى أحد مديري المؤسسات الصناعية دراسة في إحدى مصانـع الأجهـزة الالكترونية لتحديد الصلة مابين عدد الأجهزة تالفة الصنع في ثلاثين دقيقة وبين عدد الساعات التي قضاها العمال في العمل منذ بداية اليوم .

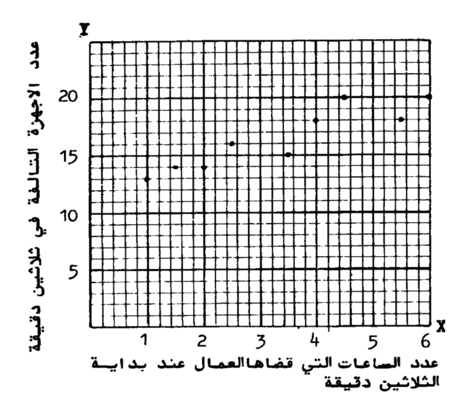
هذا وقد حصرت البيانات التي جمعت في الجدول التالي :

المطلـوب :

قياس الإرتباط ما بين عدد الساعات التي قضاها العمال في العمل (X) وعدد الأجهزة التالفة (Y) .

جدول يوضح عدد الساعــات
التي قضاها العمال وعدد الأجهزة التالفـة

عدد الساعات التي قضاها العمال في العمل عند بداية الثلاثين دقيقة	عدد الأجهزة التالفة في ثلاثين دقيقة
Number of hours, x, workmen were on the job at the start of the time period	Number defectives, y, in a 30-minute period.
1.0	13
1.5	14
2.5	16
2.0	14
3.5	15
4.5	20
4.0	18
5.5	18
6.0	20
$\Sigma X = 30.5$; $\Sigma X^2 = 128.25$; $\Sigma XY = 535.5$	$\Sigma Y = 148$; $\Sigma Y^2 = 2490$



رسم بياني يوضح العلاقة بين عدد الساعات منذ بداية العمل وعدد الاجهزة التالفة

وبالتعويض لهذه القيم يتضح لنا أن قيمة معامل الإرتباط هي

$$r = \frac{(535.5 - (30.5)(148) / 9)}{\sqrt{(128.25 - 30.5 \times 30.5 / 9)(2490 - 148 \times 148 / 9)}}$$

$$r = \frac{535.5 - 501.56}{\sqrt{(128.25 - 103.36)(2490 - 2433.78)}}$$

$$r = \frac{33.94}{\sqrt{24.89 \times 56.22}}$$

$$= \frac{33.94}{\sqrt{1399.3158}}$$

$$= \frac{33.94}{37.41}$$

$$r = 0.91$$

الخلاصــة :

يتضح من الرسم البياني ومن قيمة (r) أن هنالك ارتباط موجـب بـين عدد الساعات التي يقضيها العمال في العمل وبين عدد الأجهزة التالفة وربما كان السبب هو أنه كلما زاد عدد ساعات العمل كلما زاد الإرهاق بالنسبة للعامـل وبـذا يقـل تركيـزه وهذا بالتالي يؤدي الى هبوط كفاءة أدائه مما ينتج عنه أخطاء تؤدي الى إنتاج هذه الأجهزة التالفة .

مثـال تطبيــقي (قياس الإرتباط لبيانات مبوبــة)

الجدول أدناه يوضح الدرجات التي حصـل عليهـا مائـة طالـب في مادتـي الرياضيات والطبيعة والمطلوب حسـاب الآتي :

(أ) عدد الطلاب الذين حصلوا على الدرجات التالية 70 ـ 79 في الرياضيات و80 - 89 في الطبيعة .

(ب) نسبة الطلاب الذين حصلوا على أقل من 70 درجة في الرياضيات .

(جـ) عدد الطلاب الذين حصلوا على 70 أو أكثر في الطبيعة وأقل من 80 درجة في الرياضيات .

(د) عدد الطلاب الذين اجتازوا الامتحانين بنجاح إذا كانت أقـل درجـة لاجتياز الامتحان هي 60 .

(هـ) معامل الإرتباط بين مادتي الرياضة والطبيعـة .

الحــــل : (الحل للأجزاء أ الى د (متروك للقارىء))

(هـ) للحصول على قيمة معامل الإرتباط فإنه يتعين علينا تطبيق المعادلة رقم (26)

ولتطبيق هذه المعادلة فإنه يتعين علينا الحصول على جدول ارتباط Correla-tion table وللحصول على هذا الجدول دعنا نرمز لدرجات الرياضيات بالرمز X ولدرجات الطبيعية بالرمز Y . ثم دعنا نحسب .

$$u_x = \frac{X_i - A}{C_x} = \frac{X_i - 64.5}{10}$$

ثم نحسب

$$u_y = \frac{Y_i - B}{C_y} = \frac{Y_i - 74.5}{10}$$

جدول يوضح العلاقة بين درجة الرياضيات ودرجة الطبيعة

Mathematics Grades درجة الرياضيات

Physics Grades درجة الطبيعة	40-49	50-59	60-69	70-79	80-89	90-99	Total المجموع
90-99				2	4	4	10
80-89			1	4	6	5	16
70-79			5	10	8	1	24
60-69	1	4	9	5	2		21
50-59	3	6	6	2			17
40-49	3	5	4				12
Total المجموع	7	15	25	23	20	10	100

وبعد ذلك يمكننا الحصول على الجدول التالي مع ملاحظة أن الأرقام بداخل.

المربعات الصغيرة في كل مربع كبير هي عبارة عن حاصل ضرب u_y u_x f

		Mathematics Grades X درجة الرياضيات									
	X	44.5	54.5	64.5	74.5	84.5	94.5	f_y	$f_y u_y$	$f_y u_y^2$	مجموع أرقام المربعات الصغيرة
Y	u_y / u_x	− 2	− 1	0	1	2	3				
94.5	2				2 [ ]	4 [16]	4 [24]	10	20	40	44
84.5	1			1 [0]	4 [4]	6 [12]	5 [15]	16	16	16	31
74.5	0			5 [0]	10 [0]	8 [0]	1 [illegible]	24	0	0	0
64.5	− 1	1 [2]	4 [4]	9 [0]	5 [-5]	2 [-4]		21	− 21	21	− 3
54.5	− 2	3 [12]	6 [12]	6 [0]	2 [-4]			17	− 34	68	20
44.5	− 3	3 [18]	5 [15]	4 [0]				12	− 36	108	33
f_x		7	15	25	23	20	10	$\Sigma f_x = \Sigma f_y = N = 100$	$\Sigma f_y u_y = -55$	$\Sigma f_y U_y^2 = 253$	$\Sigma f u_x u_y = 125$
$f_x u_x$		− 14	− 15	0	23	40	30	$\Sigma f_x u_x = 64$			
$f_x u_x^2$		28	15	0	23	80	90	$\Sigma f_x u_x^2 = 236$			
مجموع ارقام المربعات الصغيرة		32	31	0	− 1	24	39	$\Sigma f u_x u_y = 125$			

(Physics Grades Y درجة الطبيعة)

قارن Check

جدول لحساب الارتباط بين درجة الرياضيات ودرجة الطبيعة

وعليه فإن :

$$S_{xy} = \left[\frac{\sum f\ u_x\ u_y - (\sum f_x\ u_x)(\sum f_y\ u_y)\ /n}{n-1} \right] C_x C_y$$

$$= \left[\frac{125 - (64)(-55)\ /\ 100}{100-1} \right] 10 \times 10$$

$$\frac{(125 - (-35.20)\ 100}{99} = \underline{161.81}$$

$$S_x^2 = \left[\frac{\sum f_x\ u_x^2 - (\sum f_x\ u_x)^2 / n}{n-1} \right] C_x^2$$

$$= \left[\frac{236 - (64)^2 / 100}{100-1} \right] 10 \times 10 = 197.01$$

$$S_y^2 = \left[\frac{\sum f_y u_y^2 - (\sum f_y u_y)^2 / n}{n-1} \right] C_y^2$$

$$= \left[\frac{253 - (-55)^2\ /\ 10}{100-1} \right] 10 \times 10 = 225.0$$

وعليه فإن معامل الارتباط r قيمته كما يـلي :

$$r = \frac{S_{xy}}{S_x S_y} = \frac{161.81}{\sqrt{197.01 \times 225.0}} = 0.7686$$

مثال تطبيقي (الإنحـــدار Regression)

أرادت إحدى الشركات معرفة مدى تأثر حجـم المبيعـات بمـا يصرف على الإعلانات في أجهزة الاعلام المقروءة والمسموعة والمرئية فقامت باختيار ست أنواع من الاعلانات كل إعلان لمدة اسبوع وأحصت حجم المبيعات لكل من هذه الأسابيع فكانت البيانات الآتية :

حجم المبيعات (Y) (بمئات الدنانير)	10.2	11.5	16.1	20.3	25.6	28.0
قيمة ما صرف على الدعاية التجارية (بمئات الدنانير) (X)	1.0	1.25	1.5	2.0	2.5	3.0

المطلــوب :

أ ـ عمل رسم بياني لهذه البيانات .

ب ـ كتابة العلاقة الخطية التي تربط ما بين حجـم المبيعـات (Y) ومـا صرف على الإعلانات (X).

جـ ـ استعمال طريقة المربعات الصغرى لتحديد هذا الخط.

د ـ حساب قيمة معامل الإرتباط بين حجم المبيعات (Y) وقيمة الاعـلان (X). هل هنالك علاقة قوية .

هـ ـ استعمل معادلة خط الانحدار التي حصلت عليها في (جـ) للتنبؤ بما سيكون عليه حجم المبيعات فيما لو صرف على الإعلانات مبلغ 220 ديناراً .

الحـــل :

أ ـ (للقـــارىء)

ب ـ ان علاقة الانحدار الخطية بين حجم المبيعات وقيمة الاعلان هي

$$Y = \alpha + \beta X + \varepsilon$$

حيث تمثـــل

α = حجم المبيعات في حالة عدم الإعلان.

β = معدل الزيادة في المبيعات لكل زيادة مقدارها مائة دينار في قيمة الإعلان.

جـ - لتحديد هذا الخط لا بد من الحصول على قيمة لتقدير α ولقيمة أخرى لتقدير قيمة β دع :

a = القيمة التقديرية لـ α

b = القيمة التقديرية لـ β

حيث:

$$a = \bar{Y} - b\bar{X}$$

$$= \frac{\Sigma Y - b\Sigma X}{n}$$

و

$$b = \frac{S_{xy}}{S_{x^2}}$$

أي أن

$$b = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma (X_i - X)^2} = \frac{\Sigma X_iY_i - \frac{(\Sigma X)(\Sigma Y)}{n}}{\Sigma X_i^2 - \frac{(\Sigma X)^2}{n}}$$

من الجدول المعطى

$\Sigma X^2 = 24.0625,$ $\Sigma Y^2 = 2346.95$

$\Sigma X = 11.25$ $\Sigma XY = 237.325$ $\Sigma Y = 111.7$

$\bar{X} = 1.875$ $\bar{Y} = 18.62$

وعليـه فـإن

$$b = \frac{237.325 - (11.25)(111.7) \; / \; 6}{24.0625 - (11.25)^2 \; / \; 6}$$

$$= \frac{237.325 - 209.4375}{24.0625 - 21.0938}$$

$$= \frac{27.8875}{2.9687} = 9.39$$

$$b = 9.39$$

أما قيمة a فتساوي

$$a = \overline{Y} - b\overline{X}$$

$$= 18.62 - 9.39 \times 1.875$$

$$= 18.62 - 17.61$$

$$a = 1.01$$

وهكذا فإن خط الإنحدار الممثل للعلاقة بين قيمة الاعلان وحجم المبيعات هـو:

$$Y = a + bX$$

$$\boxed{Y = 1.01 + 9.39X}$$

بما أن

$$S_y^2 = \frac{2346.95 - 18.62 \times 111.7}{5}$$

$$S_y^2 = 53.42$$

و

$$S_x^2 = \frac{2.9687}{5} = 0.59$$

فعليه،

$$r = \frac{b\ S_x}{S_y}$$

$$r = 9.39\sqrt{\frac{0.59}{53.42}} = 9.39 \times 0.105$$

$$r = .989$$

وهي قيمة معامل الإرتباط

هـ ـ يمكننا الآن استعمال معادلة خط الانحدار أعلاه للتنبؤ بما سيكون عليه حجم المبيعات فيما لو صرف على الاعلان مبلغ 220 دينار (لاحظ أن X تقاس بمئات الدنانير) وعليه فإن قيمة X هي $\frac{220}{100} = 2.2$

وهكذا فإن حجم المبيعات في هذه الحالة هو

$$Y = 1.01 + 9.39 \times 2.2$$

$$= 1.01 + 20.658$$

$$\boxed{Y = 21.67}$$

بمئات الدنانير

أي أن حجم المبيعات يتوقع أن يكون 2167 ديناراً.

تطبيقات خط الإنحدار على السلاسل الزمنية

إذا كان المتغير X يمثل الزمن فإن البيانات في هذه الحالة تعطي قيمة المتغير Y في أوقات مختلفة . وعليه فإنه عندما تعطى البيانات قيمة المتغير Y مع الزمن . تسمى هذه البيانات سلاسل زمنية وفي هذه الحالة فإن خط الانحدار يسمى الإتجاه العام للسلسلة الزمنية ويستعمل خط الإتجاه العام هذا لعمليات التقدير والإسقاط

والتنبؤ . هذا ولا يتسع المجال في هذا الكتاب لتناول السلاسل الزمنية بالتفصيل ولكن يهمنا الآن ونحن بصدد إيجاد الانحدار أن نحصل على خط الإتجاه العام .

مثـال تطبيـقي (السلاسل الزمنيـة)

الجدول التالي يوضح كميات الحديد التي أنتجتها احدى الدول خلال المدة من سنة 1946 الى سنة 1956 والمطلـوب :

أ ـ عمل رسم بياني لهذه البيانات .
ب ـ إيجاد خط الإتجاه العام (خط الإنحدار) .
جـ ـ تقدير الكميات المنتجة من الحديد لسنتي 1957 و 1958 ومقارنته بالكميات المنتجة الفعلية أي 112.7 و 85.3 مليون طـن .
د ـ تقدير الكميات المنتجة من الحديد لسنتي 1944 و 1945 ومقارنتة بالـكميات المنتجة فعلياً إذا علمت أنها كانت 79.7 و 89.6 مليون طن على التوالي .

السنة	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956
Production الانتاج	66.6	84.9	88.6	78.0	96.8	105.2	93.2	111.6	88.3	117.0	115.2
الزمن	0	1	2	3	4	5	6	7	8	9	10

الحـل :

أ ـ (للقـارىء)
ب ـ أن خط الإتجاه العام هو

$$Y = \alpha + \beta\ t + \varepsilon$$

حيث تمثل Y الكميات المنتجة من الحديد بملايين الأطنـان وتمثـل t الزمـن بالسنين . ولتقدير هذا الإتجاه العام يستعمل خط الإنحدار.

حيث

$$\hat{Y} = a + bt$$

$$a = \bar{Y} - b\bar{t}$$

$$b = \frac{S_{tY}}{S_t^2} = \frac{\Sigma (t_i - \bar{t})(Y_i - \bar{Y})}{\Sigma (t_i - \bar{t})^2}$$

$$= \frac{\Sigma t_i Y_i - (\Sigma t)(\Sigma Y)/n}{\Sigma t_i^2 - (\Sigma t)^2/n}$$

ويمكن الحصول على هذه القيم من الجدول كالاتي :

$\Sigma t = 55 \quad \Sigma Y = 1045.4$

$\bar{t} = 5 \quad \bar{Y} = 95.0$

$\Sigma tY = 5661.1 \quad \Sigma t^2 = 385$

$$b = \frac{5661.1 - 55 \times 1045.4/11}{385 - (55)^2/11}$$

$$b = \frac{434.1}{110} = 3.95$$

$$\boxed{b = 3.95}$$

$$a = \bar{Y} - b\bar{t} = 95 - 3.95 \times 5 = 75.2$$

$$\boxed{a = 75.2}$$

وعليه فإن خط الاتجاه العام هو:

$$\boxed{\hat{Y} = a + bt = 75.2 + 3.95t}$$

جـ ـ لتقدير كمية الحديد المنتج لسنتي 1957 و1958 فإن لسنة 1957 قيمة11 = t ولسنة 1958 قيمة 12 = t وعليه فإن القيمة التقديرية للإنتاج لسنة 1957 هي:

$$\hat{Y} = 75.2 + 3.95 \times 11 = 75.2 + 43.45$$

$$\boxed{\hat{Y} = 118.65}$$

مقارناً مع الكمية المنتجة فعلياً وهي 112.7 . أما لسنة 1958 فإن قيمة $t = 12$ وعليه فإن قيمة المنتج المقدرة هي

$$\hat{Y} = 75.2 + 39.5 \times 12 = 75.2 + 47.4$$

$$\boxed{\hat{Y} = 122.6}$$

مقارنة مع الإنتاج الفعلي وهو 85.3

ملحوظة : يلاحظ أن الكمية المقدرة لسنة 1957 أقرب الى الكمية الحقيقية فيما لو قارنا الكمية المقدرة لسنة 1958 مع الإنتاج الفعلي وهذا يوضح لنا انه في حالة التنبؤ بالمستقبل كلما بعدنا عن الفترة التي تم فيها حساب خط الإتجاه العام كلما قلت دقة التقديرات .

د ـ أما لتقدير كميات الحديد المنتجة لسنتي 1944 و1945 فإن قيمة $t = -2$ لسنة 1944 كما أنها تساوي $t = -1$ لسنة 1945 . وعليه فإن الكمية المقدرة لإنتاج سنة 1944 هي

$$\hat{Y} = 75.2 + 3.95 \times (-2)$$

$$= 75.2 - 7.9 = 67.3$$

$$\boxed{\hat{Y} = 67.3}$$

مقارنة مع الإنتاج الفعلي وهو 79.7 ولسنة 1945 فإن الكمية المقدرة للإنتاج هي

$$\hat{Y} = 75. \quad + 3.95 \times (-1)$$

$$\hat{Y} = 75.2 - 3.95 = 71.25$$

$$\boxed{\hat{Y} = 71.25}$$

مقارنة مع الكمية المنتجة الحقيقية وهي 89.6 .

ملحوظــة :

في هذه الحالة أيضاً كما كان في حالة (جـ) فإنه كلما بعدنا عن الفترة الزمنية التي حسب فيها خط الإتجاه العام كلما قلت دقة التقديرات.

وهذا ينطبق أيضاً على كل حالات التنبؤ فكلما بعدنا عن المدى الذي تتغير فيه X عند حساب خط الإنحدار كلما قلت دقة التقديرات في حالات التنبــؤ.

خطوط الإنحدار ومعامل الإرتباط الخطي :

معادلة تقدير خط الانحدار $\hat{Y} = a_0 + a_1 X$ أي معادلة خط انحدار Y على X يمكن كتابتها على الصورة :

$$Y - \bar{Y} = \frac{rS_y}{S_x}(X - \bar{X}) \quad \text{أو} \quad y = \frac{rS_y}{S_x} x \qquad (30)$$

كذلك فإن تمدير خط انحدار X على Y ، $\hat{X} = b_0 + b_1 Y$ يمكن كتابته كالآتي :

$$X - \bar{X} = \frac{rS_x}{S_y}(Y - \bar{Y}) \quad \text{أو} \quad x = \frac{rS_{xy}}{S_y} y \qquad (31)$$

ويتساوى ميل الخطوط بالمعادلات (30) ، (31) في حالة وحيدة فقط وهي إذا كانت $r \pm 1$ في مثل هذه الحالة فإن الخطين متطابقان وهناك علاقة خطية كاملة بين المتغيرين Y,X . أما إذا كانت $r = 0$ فإن الخطين متعامدان ولا يوجد ارتباط خطي بين X, Y بهذا فإن معامل الإرتباط الخطي يقيس بعد خطي الإنحدار عن بعضهما .

لاحظ أنه إذا كتبت المعادلتان (30) ، (31) كالآتي

$$\hat{X} = b_0 + b_1 Y \qquad \hat{Y} = a_0 + a_1 X$$

على الترتيب ، اذن

$$a_1 b_1 = r^2$$

ارتباط الرتـــب : Rank Correlation

بدلاً من استخدام قيم محددة للمتغيرات ، أو عندما لا يكون مثـل هذا التحديد متاحاً ، فإنه يمكن تـرتيب البيانـات حسب حجمها ، أهميتهـا، . . . أو أوغير ذلك باستخدام الأرقام n ,...,1, 2, 3. إذا رتبنا متغيرين X, Y بهذه الطريقة فإن معامل الارتباط في هذه الحالة يسمى معامل ارتباط الرتب وتعريفه كما يلي :

$$r_{rank} = 1 - \frac{6\Sigma\ D^2}{n\ (n^2 - 1)} \qquad (32)$$

حيـث :

D = الفرق بين رتب القيم المتقابلة في Y, X

n = عدد أزواج القيم (X, Y) في البيانات .

الصيغـة (٢٧) تسمـى معامـل أسبيرمـان لارتبـاط الـرتـب . Spearman Rank Correlation.

مثال تطبيقي (ارتبـاط الرتـــب) :

البيانات الآتية توضح ترتيب عشرة طلاب على ضوء نتائـج امتحـان نصف السنة وآخر السنة والمطلوب حساب معامل الارتباط لهذه الرتـــب .

نتيجة امتحان نصف السنة	8	3	9	2	7	10	4	6	1	5
نتيجة امتحان آخر السنة	9	5	10	1	8	7	3	4	2	6

الحـل : الجدول الآتي يوضح الفرق بين الرتب للعشرة طلاب.

الفرق بين الرتب D_i	-1 -2 -1 1 1- 3 1 2 1- 1-	المجموع
D^2_i	1 4 1 1 1 9 1 4 1 1	24

من الجدول يتضح أن $\Sigma D^2_i = 24$ وعليـه فإن معامل اسبيرمان للرتب

$$r_{Rank} = 1 - \frac{6 \Sigma D_i^2}{n(n^2-1)} = 1 - \frac{6\,(24)}{10\,(10^2-1\,)}$$

$$= 1 - \frac{144}{990} \quad = 1 - .145 \quad = .855$$

وهذه النتيجة تشير الى الصلة القوية بين نتائج نصف السنة وآخر السنة .

تمارين

الارتبـاط: السـؤال (1)

أ ـ أحسب قيمة معامل ارتباط اسبيرمان للمتغيرين X وهو عبارة عن الصادرات وY هو عبارة عن الواردات لدولة من الدول النامية لمدة اثني عشر شهراً.

الصادر X (ملايين الدولارات)	8	6	2	3	6	4	2	6	9	8	3	7
الوارد Y (ملايين الدولارات)	4	3	1	4	8	1	1	3	4	2	5	9

ب ـ هل هنالك ارتباط ما بين قيمة الصادرات والواردات أم لا؟

السؤال (2.)

الجـــدول أدناه يوضح أطوال وأوزان لثلاثمائة من الـرياضيين والـرياضيات والمطلوب حساب.

أ ـ عدد الذين لا يقل وزنهم عن 130 رطلاً ولكن يزيد طولهم عن 66 بوصـــة.

ب ـ الذين يزيد وزنهم عن 149 رطلاً ويزيد طولهم عن 62 بوصة.

جـ ـ عدد الذين يزيد وزنهم عن 169 رطلاً.

د ـ عدد الذين يقل طولهم عن 71 بوصـة.

هـ ـ حساب الإنحراف المعياري للأوزان.

و ـ حساب التباين للأطـوال.

ز ـ حساب التغاير بين الأطوال والأوزان.

ط ـ حساب قيمة معامل أرتباط بيرسون .

جدول يوضح الطول والوزن للرياضيين

Height X (Inches)

Weights Y (pounds) الوزن

75-78	71-74	67-70	63-66	59-62	
			1	2	90-109
	2	4	8	7	110-129
1	7	22	15	5	130-149
5	19	63	12	2	150-169
12	32	28	7		170-189
7	20	10	2		190-209
2	4	1			210-229

السؤال (3)

المطلوب حساب معامل ارتباط الرتب (معامل ارتباط اسبيرمان) من الجدول أدناه الذي يوضح أطوال الآباء وأبنائهم .

جدول يوضّح طول الآباء والأبناء

طول الأب (X) بالبوصات	65	63	67	64	68	62	70	66	68	67	69	71
طول الابن (Y) بالبوصات	68	66	68	65	69	66	68	65	71	67	68	70

الانحدار : السؤال (1)

الجدول التالي يوضح متوسط دخل الفرد لتسعة من الدول (X) ومتوسط ما صرف على الطالب الواحد للتعليم العام (Y)

جدول يوضح متوسط دخل الفرد ومتوسط الصرف على التعليم

الدولة	متوسط دخل الفرد X (بالدولار)	متوسط ما صرف على الطالب في التعليم العام Y (بالدولار)
1	2520	535
2	4272	922
3	3568	695
4	3944	842
5	2194	476
6	2890	609
7	4421	1237
8	3779	904
9	3051	657

المطلوب:

أ- رسم هذه البيانات بحيث يتدرج المحور الأفقي من صفر إلى 5000 ويتدرج المحور الرأسي من صفر الى 1400 .

ب- رسم خط تقديري لوصف العلاقة بين المتغيرين Y,X

جـ- حساب خط الانحدار .

$$\hat{Y} = a + bX$$

بطريقة المربعات الصغرى وقارن ما بينه وبين الخط في الجزء (ب)

السـؤال (2)

الجدول أدناه يوضح الإنتاج السنوي للسجاير في إحدى الدول الصناعيـــة .

جدول الانتاج السنوي للسجائر

السنة	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954
عدد السجاير المنتجة بالمليون	98,2	92,3	80	89,1	83,5	68,9	69,2	67,1	58,3	81,7

المطـلوب :

أ ـ رسم البيانات رسماً بيانياً .

ب ـ ايجاد معادلة خط الانحدار .

جـ ـ قدر إنتاج السجاير لسنة 1955 .

د ـ قدر الانتاج لسنة 1948 وقارنه مع المنتج الفعلي .

الفصل الخامس

الأرقــام القياسيــة

Index Numbers

الفصل الخامس

الأرقام القياسية

Index Numbers

الرقم القياسي هو مقياس إحصائي مصمم لقياس التغيرات التي تحـدث في متغير ما أو لمجموعة من المتغيرات خلال زمن محدد أو منطقة محددة الى زمن آخر أو مكان آخر.

يهتم الباحثون في الوقت الحالي بدراسة التغير الذي يطرأ على الظواهر أي كان نوعها ، وهـم في ذلك يستخدمـو ن الكثـير من المقـاييس لتحـديد هذا التغـير بل دراسته ، ومهما تعددت المقـاييس المستخدمـة لتحـديد هذا التغـير إلا أن الأرقـام القياسية Index Numbers هي أكثر هذه المقاييس انتشاراً.

وكلمة Index من وجهة النظـر الإحصـائية تعنـي المؤشر الإحصائـي الـذي يستخدم في قياس التغير الذي يطرأ على العديد من الظواهر الإجتماعية والإقتصادية ، وخصوصاً تلك الظواهر التي تقع تحت تأثير عوامل مختلفة فتتغير من وقت الى آخر أو من مكان إلى آخر.

وعلى ذلك فإن الرقم القياسي يمكن أن يقيس لنا التغير في حجم الإنتاج لنوع معين من السلع أو لمجموعة معينة من السلع ذات النوعية الواحدة أو التغير في أسعار أو كميات سلع التصدير أو الإستيراد أو التغير في تكلفة إنتاج الوحدة من إنتاج معين أو قياس التغير في حجم السكان وكذلك التغـير في أجـور العمال في الصناعـات المختلفة والتي تتغير من مكان إلى مكان ومن وقت إلى آخر.

ومن الأمـور الهامـة عنـد تركيب الرقـم القياسي ، اختيار فتـرة (أو مكان) الأساس التي نعتبرها لتركيب الرقم . وفي العادة تكون فترة الأساس سابقة للفترة التي نريد مقارنتها . وعند اختيار فترة أو مكان الأساس ، والتي تنسب اليها أرقام سنة المقارنة ، يجب أن تكون مختارة عند فترة متميزة بالإستقرار الإقتصادي وخالية من العوامل الشاذة التي تؤثر على هذه الأرقام مثل الحروب ولا تكون بعيدة جداً عن سنوات المقارنة .

وللأرقام القياسية تطبيقات عديدة في مجالات مختلفة فلم تعد وسيلة في أيدي الإقتصـاديين في دراساتهـم التحليلية لقياس التغـيرات التـي تطـرأ على الظواهـــر الإقتصادية فحسب ، بل أصبحت في وقتنا هذا وسيلة في أيدي المهتمـين بالعلوم الإجتماعية والنفسية والإدارية والعلوم الزراعية لعمل المقارنات وقياس التغيرات في هذه المجالات المختلفة .

وتوجد أرقام قياسية عديدة تستعمل في مختلف الميادين مثـل الرقـم القياسي لأسعار الجملة وأسعار التجزئة وأسعار مجموعات خاصة من السلـع مثـل مجموعـة الوقود ومجموعة المواد الخام ومجموعة السلع الوسيطة ومجموعة السلـع الإسـتثمارية ومجموعة السلع الإستهلاكية المعمرة وغير المعمرة ، كلها مجموعات رئيسية تدخل في تركيب الرقم القياسي للواردات ، وكذلك توجـد أرقـام قياسية للإنتـاج الزراعـي والإنتاج الصناعي والأجور وتكلفة المعيشة .

الأرقام القياسية البسيطة Unweighted Index Numbers

1 ـ مناسيب الأسعار Price relatives

من أبسط الأمثلة للرقم القياسي هو منسوب السعر ، وهو نسبة السعر لسلعة واحدة في فترة المقارنة الى سعرها في فترة أخرى تسمى بفترة الأساس أو فترة الاسناد . وللتسهيل سوف نفترض أن الأسعار ثابتة لأي فترة . فإذا لم يكن هذا صحيحاً فإنه يمكن استخدام متوسط ملائم حتى نجعل هذا الفرض صحيحاً .

إذا كانت P_0 تمثل سعر السلعة خلال فترة الأساس و P_n سعرها خلال فترة المقارنة ، فإنه بالتعريف .

$$\text{منسوب السعر} = \frac{P_n}{P_0} \qquad (1)$$

ويعبر عنه بشكل عام في صورة نسبة مئوية بضربه في 100 وبشكل أكثر عمومية إذا كانت P_a , P_b هي أسعار سلعة خلال الفترات a, b على الترتيب ، فإن منسوب السعر في الفترة b بالنسبة للفترة a يعرف بأنه P_b / P_a ويرمز له بالرمز $P_{a/b}$ وسنجد هذا الرمز مفيد فيما بعد . لهذا فإن منسوب السعر في المعادلة (1) يمكن أن يرمز له بالرمز $P_{o/n}$.

مثـال (1) :

إفترض أن أسعار المستهلكين لعنصر معين في السنوات 1955, 1960 هي 25, 30 قرشاً على الترتيب : فإذا أخذنا 1955 كسنة أساس و1960 سنة المقارنة ، فإن

$$\text{منسوب السعر} \qquad P_{1955/1960} = \frac{\text{السعر في } 1960}{\text{السعر في } 1955}$$

$$= \frac{30}{25} = \frac{6}{5}$$

$$= 120\%$$

أو باختصار 120 بحذف علامة % كما هو متبع غالباً في المؤلفات الاحصائية. هذه النتيجة تعني ببساطة أن سعر العنصر سنة 1960 أصبح 120% من سعره في سنة 1955 أي زاد بنسبة 20% .

مثـال (2) :

بأخذ 1960 كسنة أساس و 1955 هي سنة المقارنة في المثال الأول فإن .

$$\text{منسوب السعر} = \frac{P_n}{P_0} = \frac{25}{30} = {}^5/_6 = 83^1/_3\%$$

أو باختصار $83^1/_3$ وهذا يعني أنه في 1955 كان سعر العنصر هو $83^1/_3\%$ من سعره في 1960 أي أنه كان ينقص بنسبة $16^2/_3\%$.

لاحظ أن منسوب السعر لفترة معينة بالنسبة لنفس الفترة سيكون دائماً 100% أو 100 وعلى وجه الخصوص فإن منسوب السعر المقابل لفترة الأساس يصبح دائماً 100 وهذا يوضح الرمز الذي يستخدم غالباً في المؤلفات الإحصائية بكتابة 100 = 1955 للإشارة الى أن سنة 1955 أخذت كسنة أساس.

خصائص مناسيب الأسعار Price Relatives Properties

أ ـ خاصية التطابق

وهذه تقرر ببساطة أن منسوب السعر لفترة معينة بالنسبة لنفس الفترة يساوي 1 أو 100%.

ب ـ خاصية الإنعكاس في الزمن Time Reversal Test

وهذه تقرر أنه إذا أحللنا فترتين كل محل الأخرى، فإن مناسيب الأسعار المقابلة تكون كل منها مقلوب الأخرى.

وهناك أنواع أخرى من الإختبارات أو الخصائص. والرقم القياسي الذي يحقق الإختبار يعتبر هو المفضل Desirable.

2 ـ مناسيب الكمية أو الحجم Quantity Relatives

بدلاً من مقارنة أسعار السلعة قد نهتم بمقارنة كميات أو مجموعة السلع مثل كمية أو حجم الإنتاج، الإستهلاك، التصدير، وغيرها. في مثل هذه الحالات نتكلم عن مناسيب الكمية كما في حالة الأسعار، إذا كانت q_0 تعبر عن كمية السلعة

المنتجة ، المستهلكة ، المصدرة وغير ذلك خلال فترة الأساس ، بينما q_n تعبر عن كمية الإنتاج ، الإستهلاك وغير ذلك المقابلة خلال فترة المقارنة ، فإننا نعرف .

$$\frac{q_n}{q_o} = \text{منسوب الكمية أو الحجم} \qquad (2)$$

ويعبر عنها بصفة عامة في شكل نسب مئوية .

3 ـ مناسيب القيمة Value Relatives

إذا كانت P هي سعر السلعة خلال فترة ما و q هي الكمية المنتجة أو الحجم المباع ، المنتج وغير ذلك ، خلال الفترة إذن pq تسمى القيمة الإجمالية . بهـذا فإن بيعت 1000 وحدة بسعر 30 قرشاً لكل وحدة فإن القيمة الإجمالية هي
$(30)\ (1000) = 30000$ قرشاً

إذا كانت P_o تعبر عن السعر و q_o عن الكمية لسلعة خلال فترة الأساس بينما p_n تعبر عن السعر المقابل و q_n عن الكمية المقابلة خلال الفترة المعطاة ، كذلك فإن القيمة الإجمالية خلال هذه الفترات هي V_o لفترة الأساس و V_n للفتـرة المعطـاة فإننـا نعرف :

$$V_{a/b} = P_{a/b}\, q_{a/b} \qquad (3)$$

4 ـ الطريقة التجميعية البسيطة Simple Aggregate Index

في هذه الطريقة لحساب الرقم القياسي للأسعار ، فإننا نعبر عن مجموع أسعار السلع في سنة المقارنة كنسبة مئوية من مجموع أسعارها في سنة الأساس .

$$\frac{\Sigma\, P_n}{\Sigma\, p_o} = \text{الرقم القياسي التجميعي البسيط} \qquad (4)$$

حيث ΣP_0 = مجموع أسعار السلع في سنة الأساس.
ΣP_n = المجموع المقابل لأسعار السلع في سنة المقارنة.

حيث يعبر عن النتيجة كنسبة مئوية كما هو بالنسبة للأرقام القياسية بشكل عام، وعلى الرغم من أن هذه الطريقة لها الميزة بأنها سهلة التطبيق، إلا أن لها عيبين كبيرين يجعل استخدامها غير مستحب: -

أولاً: لا تأخذ في الإعتبار الأهمية النسبية للسلع المختلفة فمثلاً طبقاً لهذه الطريقة، فإن أوزاناً متساوية تعني أن نفس الأهمية سوف تعطى للألبان ولمعجون الحلاقة عند حساب الرقم القياسي لتكلفة المعيشة.

ثانياً: الوحدات المستخدمة في تمييز السعر، مثل الجرام، وغيرها تؤثر على قيمة الرقم القياسي.

5 - الوسط الحسابي لمناسيب الأسعار Arithmatic Mean of Price Relatives

هناك عديد من الصيغ تعتمد على الطريقة المستخدمة في الحصول على أوساط مناسيب الأسعار، مثل الوسط الحسابي، الوسط الهندسي، الوسط التوافقي، الوسيط، وما إلى ذلك. فإذا استخدمنا الوسط الحسابي، على سبيل المثال فإننا نحصل على

الوسط الحسابي البسيط للرقم القياسي لمناسيب الأسعار $= \frac{\Sigma P_n / P_0}{N}$ (5)

حيث $\Sigma P_n / P_0$ = مجموع مناسيب أسعار جميع السلع.
N = عدد مناسيب أسعار السلع المستخدمة (عدد السلع.)

وعلى الرغم من أن هذه الطريقة تتخلص من العيب الثاني الموجود في الطريقة التجميعية البسيطة يظل العيب الأول موجوداً بها.

الطريقة التجميعية المرجحة Weighted Index Numbers

للتغلب على عيوب الطريقة التجميعية البسيطة ، ترجح أسعار كل سلعة باستخدام معامل ملائم ويستخدم غالباً كمية أو حجم السلعة المباعة خلال فترة الأساس ، أو سنة المقارنة أو سنة نموذجية (والتي قد تتضمن متوسط عدد من السنوات). هذه الأوزان تشير الى أهمية السلعة المعنية . وهناك ثلاث صيغ ممكنة تعتمد علي ما إذا كنا سنستخدم كميات سنة الأساس أو سنة المقارنة أو سنة نموذجية ونعبر عنها بالرموز q_t, q_n, q_o على الترتيب .

1 ـ رقم لاسبيرز القياسي أو طريقة سنة الأساسي : Laspeyres' Index

الرقم القياسي التجميعي المرجح باستخدام كميات سنة الأساس

$$I_L = \frac{\Sigma P_n q_o}{\Sigma P_o q_o} \qquad (6)$$

2 ـ رقم باشي القياسي (أو طريقة سنة المقارنة) Paache's Index

الرقم القياسي التجميعي المرجح باستخدام كميات سنة المقارنة

$$I_P = \frac{\Sigma p_n q_n}{\Sigma p_o q_n} \qquad (7)$$

3 ـ طريقة النسبة النموذجية :

إذا اعتبرنا q_t تعبر عن وزن الكمية خلال فترة نموذجية فإننا نعرف الرقم القياسي التجميعي المرجح باستخدام كميات السنة النموذجية

$$I = \frac{\Sigma p_n q_t}{\Sigma p_o q_t} \qquad (8)$$

عندما تكون $t = n, t = 0$ فإن هذه الصيغة تؤول الى الصيغة (6) والصيغة (7) على الترتيـــب .

رقم فيشر المثالي : Fisher's Ideal Index Number

نعرف الرقم القياسي المثالي لفيشر

$$I_F = \sqrt{\left[\frac{\Sigma p_n q_0}{\Sigma p_0 q_0}\right] \left[\frac{\Sigma P_n q_n}{\Sigma p_0 q_n}\right]} \quad (9)$$

وهذا الرقم القياسي عبارة عن الوسط الهنـدسي للرقمـين القياسيين لكل من لاسبيرزوباشي الموضحين بـالمعادلتـين (6) و (7) ورقم فيشر المثـالي يحقق كلا من اختباري الإنعكاس في الزمن والإنعكاس في المعامل ، وهذا ما يعطيه بعض المزايا النظرية عن الأرقام القياسية الأخرى .

رقم مارشال ـ ادجورث القياسي : Marshall- Edgeworth Index

يستخدم رقم مارشال ادجورث القياسي الصيغة التجميعية المرجحة باستخدام طريقة السنة النموذجية حيث الأوزان هي الوسط الحسابي لكميات سنة الأســاس وكميات سنة المقارنة .

أي $q_t = {}^1/_2 (q_0 + q_n)$ وبالتعويض بهذه القيمة لـq_t في المعادلـة (8) ، نحصـــل على : ـ

رقم مارشال ـ ادجورت القياسي للأسعـــار

$$I_{M\text{-}E} = \frac{\Sigma p_n (q_0 + q_n)}{\Sigma p_0 (q_0 + q_n)} = \quad (10)$$

الوسـط المرجح للمناسيب :

للتغلب على العيوب في طريقة الوسط البسيط للمناسيب يمكن أن نستخدم

متوسطاً مرجحاً للمناسيب . والوسط المرجح الأكثر شيوعاً في هذا المجال هو الوسط الحسابي المرجح ، على الرغم من أنه يمكن استخدام أوساطاً مرجحة أخرى مثل الوسط الهندسي المرجح .

في هذه الطريقة نرجح كل منسوب سعر بالقيمة الإجمالية للسلعة وذلك بدلاً من الوحدات النقدية مثل الدولار . . . الخ وبما أن قيمة السلعة نحصل عليها بضرب السعر p للسلعة في الكمية q فإن الأوزان تعطى بالصيغة pq وهناك ثلاثة صيغ يمكن استخدامها وهذه تعتمد على ما إذا كنا نستخدم قيمة سنة الأساس أو سنة المقارنة أو سنة نموذجية ويعبر عن ذلك بالقيم p_oq_o , p_nq_n , p_tq_t على الترتيب .

أ ــ الوسط الحسابي المرجح لمناسيب الأسعار باستخدام قيمة سنة الأساس كأوزان :

$$\frac{\Sigma (p_n/p_o)(p_oq_o)}{\Sigma p_oq_o} = \frac{\Sigma p_nq_o}{\Sigma p_oq_o} \qquad (11)$$

ب ــ الوسط الحسابي المرجح لمناسيب الأسعار باستخدام قيمة سنة المقارنة كأوزان :

$$\frac{\Sigma (p_n / p_o)(p_nq_n)}{\Sigma p_nq_n} \qquad (12)$$

جـ ــ الوسط الحسابي المرجح لمناسيب الأسعار باستخدام قيمة سنة نموذجية كأوزان :

$$\frac{\Sigma (p_n / p_o)(p_tq_t)}{\Sigma p_tq_t} \qquad (13)$$

لاحظ أن المعادلة (11) تعطي نفس صيغة لاسبيرز المعرفة بالمعادلة (6) .

الأرقام القياسيــة المرجحــة للكمية أو القيمة :

الصيغ الموضحة أعلاه التي تعرف بالأرقام القياسية للأسعار يمكن بسهولـة تعديلها للحصول على الأرقام القياسية للكمية أو القيمة وذلك ببساطة بإبدال q, p كل محل الآخر على سبيل المثال بابدال P بـ q في المعادلة (5) نحصل على .

$$\frac{\Sigma\, q_n / q_0}{N} \qquad (14)$$

حيث $\Sigma\, q_n / q_0$ = مجموع مناسيب كميات جميع السلع

N = عدد مناسيب أسعار جميع السلع المستخدمــة

الرقم القياسي التجميعي المرجح باستخدام اسعار سنة الاســاس كاوزان

الرقم القياسي

$$I = \frac{\Sigma\, q_n p_0}{\Sigma\, q_0 p_0} \qquad (15)$$

يسمى برقم لاسبيرزالقياسي للكميات ونحصل عليه بالابدال في المعادلة (6) .

الرقم القياسي التجميعي المرجح باستخدام أسعار سنة المقارنة كأوزان

الرقم القياسي :

$$I = \frac{\Sigma\, q_n p_n}{\Sigma\, q_0 p_n} \qquad (16)$$

يسمى برقم باشي القياسي للكمياتونحصل عليه بالابدال في المعادلة (6)

ويمكن باتباع الخطوات في المعادلات من (8) الى (13) ن نحصل على صيغة للرقم القياسي للكميات .

. الرقم القياسي للقيم

الرقم القياسي للقيمــة = $\frac{\Sigma\, q_n p_n}{\Sigma\, q_0 P_0}$ (17)

حيث $\Sigma\ p_0q_0$ = القيمة الإجمالية لجميع السلع في فترة الأساس .

$\Sigma\ p_nq_n$ = القيمة الإجمالية لجميع السلع في فترة المقارنة .

وهذا رقم قياسي تجميعي بسيط ، حيث أن القيمة لم ترجح .

ويمكن وضع صيغ أخرى حيث نستخدم الأوزان للدلالة على الأهمية النسبية للعناصر .

استعمالات الارقام القياسية

تستعمل الارقام القياسية في كثير من المجالات الاقتصادية وسنتعرض فيما يلي لثلاث من هذه الاستعمالات .

1 ـ الرقم القياسي لتكاليف المعيشة Consumer Price Index

إذا أردنا أن نخرج من نطاق البحث في قياس مدى الإرتفاع أو الإنخفاض في سعر سلعة معينة الى محاولة قياس مدى التغير في مجموع السلع والخدمات التي يستهلكها أفراد المجتمع سنوياً ، وأن نعبر عن هذا كله برقم واحد . . ماذا نصنع في هذه الحالة؟

من الواضح أن محاولة البحث في الأرقام القياسية لألوف السلع التي يستهلكها المجتمع ، كمقدمة لإيجاد رقم قياسي لتكاليف المعيشة ، سوف يعني بذل جهود مضنية تستغرق زمناً طويلاً ، وهو أمر غير عملي . ومن الناحية العملية يلجأ الإحصائيون في معظم البلدان الى اختيار نحو 300 سلعة فقط من مجموع السلع التي يشتريها الناس ، ويستخدمون هذه السلع كأساس لإيجاد الرقم القياسي لتكاليف المعيشة . ولا تختار هذه السلع عشوائياً من مجتمع السلع ، وإنما تختار على أساس الأهمية النسبية لكل سلعة وعلى أساس توفر بيانات دقيقة عن أسعارها .

ويقتضي هذا بالطبع أن تكون هناك دراسات لميزانية الأسرة في مناطق مختلفة من البلاد حتى تختار الثلاثمائة سلعة على أساس موضوعي سليم ، بحيث تكون السلع المختارة هي أهم سلع ينفق الناس لاستهلاكها .في كثير من الأحيان نلاحظ أن

التغير في أسعار السلع المختارة قد يكون أيضاً معبراً عن التغير في أسعار سلع لم تختار في العينة.

كذلك يجري البحث في هذا الموضوع على أساس اختيار عينة من المدن بحيث تكون هذه المدن ممثلة للمجتمع ككل (يراعى أن العادات الإستهلاكية للناس قد تتغير من مدينة لمدينة لتغيرات ظروف المناخ فيها وخصوصاً في البلاد الواسعة ويحسب الرقم القياسي السعري للمستهلك في كل مدينة على حدة، ثم يحسب الرقم القياسي الإجمالي بأخذ متوسط هذه الأرقام القياسية بعد مراعاة الأوزان المختلفة لحجم السكان في كل مدينة.

ورغم كل التبسيطات التي سبق ذكرها (عينة من السلع،عينة من المدن... الخ) فإن تركيب الرقم القياسي للإستهلاك هو عمل شديد الصعوبة ومليء بالمشاكل المعقدة.

فجمع البيانات المناسبة عن السلع الأكثر أهمية في استهلاك الناس ليس بالعمل البسيط. وقد تكون مسألة جمع بيانات الأسعار سهلة نسبياً، ولكن المشكلة التي تواجه المشتغلين في هذا الميدان هو جعل الأسعار قابلة للمقارنة من شهر الى شهر ومن مدينة الى مدينة. فحتى سلعة بسيطة كالخبز قد تجدها مختلفة الأحجام من مدينة الى مدينة، وقد تباع أرغفة من أحجام مختلفة بنفس المبلغ من النقود. ولا يعني هذا بالطبع أن أسعارها واحدة. فإذا دخلنا في ميدان الملابس فسوف نجد أن قابلية المقارنة بين السلع المسعرة تصبح أكثر صعوبة، ونفس الشيء ينطبق على الإسكان إذ من الصعب مقارنة المساكن فليس من السهل أن نقول عن سكنين أنهما متساويان من ناحية التسهيلات والراحة والمتعة.

لكل هذا من الصعب إيجاد رقم قياسي لتكاليف المعيشة يرضى عنه الجميع ولكن الخبرة المستمرة في هذا الميدان كفيلة بان تحسن أساليب تركيب هذا الرقم سنة بعد أخرى.

2 ـ الرقم القياسي لأسعار الجملة : Wholesale Price Index

الرقم القياسي لأسعار الجملة لا يعني قياس التغير في أسعار السلع حسب

أهمية تداولها التجاري . لذلك فكلمة الجملة Wholesale لا تشير الى أسعار تاجر الجملة أو أسعار الموزعين ولكنها تشير الى أسعار حصص كبيرة من السلع في سوقها الأولى Primary Market

والأسعار المستخدمة في هذا المقياس هي أسعار السلع المتبادلة في حصص منظمة أو أسواق منتظمة أو أسعار المصنع أو المنتج .

والتغير في الأسعار المقاس هنا يكون سعر سلعة واحدة أو مجموعة سلع أو خليط من أسعار سلع وهي تكون عادة أسعار حقيقية للسلع لا تتأثر بالتغيرات في الكمية أو النوعية ولا بالتوزيع أووحدةالاسعار.

مثل الرقم القياسي لتكاليف المعيشة يقتضي الرقم القياسي لأسعار الجملة إجراء مسوحات بالعينة.ولأهمية حركة أسعار الجملة في الاقتصاد يلجأ الإحصائيون الى أخذ 2 600 سلعة (الفين وستمائة) وعادة تختار أسعار السلع في منتصف الشهر ، وهذه السلع لا تختار عشوائياً من سلع تجارة الجملة بل تختار عن طريق أخصائيين حسب تصنيفات هذه السلع .

وفي الوقت الحاضر تحسب الأسعار القياسية منفصلة للسلع حسب تصنيفاتها . ومن أهم ذلك التصنيف حسب المراحل العملية للسلعة على سبيل المثال سلع المواد الخام التي تتداول في السوق،سلع المواد الوسيطة والسلع النهائية .

وهناك تصنيفات أخرى،فنجد مثلاً إنتاج الحقول الزراعية مقسم على عدد من التصنيفات هي السلع الطازجة من الخضروات والفواكه وغير الطازجة والحبوب ، والدواجن ، والثروة الحيوانية وهناك أيضاً تصنيف الإنتاج الصناعي .

ويحسب الرقم القياسي للأسعار بطريقة الوسط الحسابي المرجح لمناسيب الأسعار باستخدام قيمة سنة المقارنة كأوزان (أنظر المعادلة 12) .

الرقم القياسي للإنتاج الصناعي Index of Industrial Production

هذا الرقم يقيس التغيرات المادية في حجم أو كمية مخرجات مصانع دولة معينة ، كمخرجات المناجم والكهرباء والغاز .

بدأ تطوير هذا النوع من الأرقام القياسية في الأعوام 1912 و1922 واستخدم بنجاح في عام 1927 وكان يحتوي على ستون سلسلة من البضائع المصنعة ومنتجات المناجم . ويمثل هذا الرقم أما بطريقة مباشرة أو غير مباشرة حوالي 80% من إنتاج الدولة من هذه البضائع والمنتجات .

وتمشياً مع التوسع الإقتصادي أدخلت تعديلات في هذا الرقم في الأعوام 1940 و1953 و1959 وهدفت هذه التعديلات الى أن يصبح هذا الرقم القياسي أكثر شمولية ولكي يكون قابلاً للمقارنة مع الأرقام القياسية للإنتاج في بلدان أخرى كما أضيفت منتجات الكهرباء والماء والغاز .

وفي عام 1971 كان آخر تعديل في هذا الرقم حيث مثلت مباشرة مواد التركيب والإنشاء. ولكن توزيع المنتجات الصناعية واستعمالاتها في الإنشاء ما زالت تغطى بصورة غير مباشرة ، أما الإنتاج في الصناعة الإنشائية نفسها والإنتاج الزراعي وإنتاج وسائل النقل ومختلف الصناعات التجارية والخدمية فهي مازالت غير مشمولة في هذا الرقم .

يتم حساب الرقم القياسي للإنتاج بعد تقسيم القطاعات المختلفة في شكل مجموعات والتي بدورها تؤخذ كمؤشرات لـ 227 سلسلة فردية من التسجيلات الشهرية للإنتاج وتجمع في شكل رقم واحد كلي ، فمثلاً واحد من هذه التقسيمات التجميع الصناعي . كأن تحصر جميع الصناعات ، المناجم والصناعات الخدمية وصناعات أخرى وهذه الصناعات بدورها تقسم إلى شبه مجموعات كالصناعات المعمرة والصناعات الغير معمرة وهكذا الى أن يشمل الرقم كل القطاعات في الإقتصاد .

وحساب الرقم الكلي يقود الى الرقم التجميعي للأوزان الثابتة وهو نسبة القيمة المضافة في ذلك الشهر أو الفترة الى سنة الأساس .

$$I = \frac{\Sigma\, q_o p_o}{\Sigma\, q_n p_o} \cdot 100$$

ويختلف رقم الإنتاج الصناعي عن أرقام المستهلك والجملة في أن القصد منه هو قياس الوضع الحالي للإنتاج أو التجارة وهذا المؤشر الحساس الشامل للتجارة يزود بشكل أساسي المؤسسات والهيئات التي تتأثر نشاطاتها الحالية وخططها المستقبلية بالتغيرات المستمرة في التجارة والانتاج .

أمثلة تطبيقية (الأرقام القياسية)

الطريقة التجميعية البسيطة

سؤال (1) ـ الجدول أدناه يوضح متوسط أسعار الجملة لأسعار منتجات الحليب للسنين 1949 ، 1950 ، 1958 ـ المطلوب حساب الرقم القياسي التجميعي البسيط لهذه المنتجات لسنة 1958 :

(أ) مستعملاً سنة 1949 كسنة الأساس .

(ب) مستعملاً متوسط سنتي 1949 و 1950 كفترة الأساس .

الاسعار (فلس للكيلو)

المنتج	1949	1950	1958
الحليب	395	389	413
الزبد	615	622	597
الجبن	348	354	389

الحــل :

أ ـ الرقم القياسي التجميعي البسيط =

$$= \frac{\text{مجموع الأسعار في سنة } 1958}{\text{مجموع الاسعار في سنة } 1949}$$

$$= \frac{413 + 597 + 389}{395 + 615 + 348} = \frac{1399}{1358} = 103\ \%$$

الخلاصة: إن متوسط أسعار الجملة في 1958 يساوي % 103 من متوسط أسعار الجملة في 1949 (أي بزيادة قدرها % 3) .

ب ـ لاستعمال أسعار سنتي 1950 & 1949 كسنة الأساس لا بد لنا من الحصول على متوسط الأسعار لكل منتج على حدة في هاتين السنتين.

متوسط أسعار الحليب للسنتين 1950—1949 =

$$\frac{1}{2}(395 + 389) = 392$$

متوسط أسعار الزبد للسنتين 1950—1949 =

$$\frac{1}{2}(615 + 622) = 618.5$$

متوسط أسعار الجبن للسنتين 1950—1949 =

$$\frac{1}{2}(348 + 354) = 351$$

$$\text{الرقم القياسي التجميعي البسيط} = \frac{\text{مجموع الاسعار في سنة } 1958}{\text{مجموع الاسعار في فترة الأساس } (1949—1950)}$$

$$= \frac{413 + 597 + 389}{392 + 618.5 + 351} = \frac{1399}{1361.5} = 102.8\ \%$$

الخلاصة : إن متوسط أسعار منتجات الحليب في سنة 1958 قد زاد بنسبة % 2,8 عنه في سنتي 1950—1949

عيـوب هذه الطريقـة :إنها لا تأخذ في الحسبان الكميات المستعملة من كل منتـج .

سؤال (2) بالرجوع الى الجدول للسؤال رقم (1) المطلوب حساب الوسط البسيط لمناسيب الأسعار .

الحـل :

منسوب أسعار الحليب $= \dfrac{\text{سعر الحليب في سنة 1958}}{\text{سعر الحليب في سنة 1949}} = \dfrac{413}{395}$

$= 104.6\ \%$

منسوب أسعار الزبد $= \dfrac{\text{سعر الزبد في سنة}}{\text{سعر الزبد في سنة 1949}} = \dfrac{597}{615}$

$= 97.1\ \%$

منسوب أسعار الجبن $= \dfrac{\text{سعر الجبن في سنة 1958}}{\text{سعر الجبن في سنة 1949}} = \dfrac{389}{348}$

$= 111.8\ \%$

الوسط الحسابي لمناسيـب الأسعار $= \dfrac{104.6 + 97.1 + 111.8}{3}$

$= \dfrac{313.5}{3} = 104.5$

الطريقة التجميعية المرجحـة

سؤال (3): الجدول التالي يوضح الكميات المنتجة من منتجـات الحليب للسنـين 1949 ، 1950 ، 1958 .

الكميات المنتجة
(بملايين الكيلو جرامات)

المنتج	1958	1950	1949
الحليب	10.436	9717	9675
الزبد	115.5	115.5	117.7
الجبن	82.79	74.399	77.93

علماً بأن أسعار هذه المنتجات قد أعطيت في الجدول المرافق للسؤال الأول.

المطلوب حساب الآتي:

أ ـ رقم لاسبيرز القياسي (سنة الأساس 1949).

ب ـ رقم باشي القياسي (سنة الأساس 1958).

جـ ـ رقم فيشر القياسي المثالي.

د ـ رقم مارشال ـ ادجورث القياسي.

الحل:

كل هذه الطرق الثلاث تلتقي في أنها ترجح الرقم القياسي باستعمال الكميات المنتجة

أ ـ رقم لاسبيرز القياسي (سنة الأساس 1949)

$$I_L = \frac{\Sigma\, q_n\, p_0}{\Sigma\, q_0\, p_0} \quad =$$

$$= \frac{\text{مجموع (أسعار 1958) (الكميات المنتجة 1949)}}{\text{مجموع (أسعار 1949) (الكميات المنتجة 1949)}}$$

$$= \frac{((413 \times 9675) + (597 \times 117.7) + (389 \times 77.93)}{((395 \times 9675) + (615 \times 117.7) + (348 \times 77.93)}$$

$$= \frac{4096356.67}{3921130.14} = 104.5\ \%$$

ب ـ رقم باشي القياسي (سنة الأساس 1958)

$$= I_P = \frac{\Sigma p_n\ q_n}{\Sigma\ P_0 q_n}$$

$$= \frac{\text{مجموع (اسعار 1958) (الكميات المنتجة 1958)}}{\text{مجموع (أسعار 1949) (الكميات المنتجة 1958)}}$$

$$= \frac{(413 \times 10436) + (597 \times 115.5) + (389 \times 82.79)}{((395 \times 10436) + (615 \times 115.5) + (348 \times 82.79)}$$

$$= \frac{4411226\ ,81}{4222063\ ,42} = 104.4\%$$

جـ ـ رقم فيشر القياسي الأمثل

$$I_F = \sqrt{\frac{\Sigma\ p_n q_n}{\Sigma\ p_0\, q_n} \quad \frac{\Sigma\ p_n q_0}{\Sigma\ p_0\, q_0}}$$

= الوسط الهندسي لرقم لاسبيرز I_L ورقم باشي I_p القياسيين

$$= I_F = \sqrt{I_P I_L}$$

$$= \sqrt{104\ 5 \times 104.4}$$

$$= \sqrt{10920.24}$$

$$= 104.5$$

$$M\text{-}E = \frac{\Sigma p_n (q_o + q_n)}{\Sigma p_o (q_o + q_n)}$$

د ـ رقم مارشال ـ ادجورث القياسي

$$= \frac{413(9675+10436)+597(117.7+115.5)+389(77.93+82.79)}{395(9675+10436)+615(117.7+115.5)+348(77.93+82.79)}$$

$$= \frac{413\ (20111)\ +\ 597\ (233.2)\ +\ 389\ (160.72)}{395\ (20111)\ +\ 615\ (233.2)\ +\ 348\ (160.72)}$$

$$= \frac{8507583.48}{8143193.56} = 140.5\%$$

الوسط البسيط للمناسيب

سؤال

استعمل طريقة الوسط البسيط لمناسيب الأسعار لحساب الرقم القياسي لأسعار منتجات الحليب (كما في الجدول للمثال نمرة (1)) لسنة 1958 مستعملاً سنتي 1949 ـ 1950 كفترة اساس .

الحل

منسوب أسعار الحليب $= \dfrac{\text{سعر الحليب في 1958}}{\text{سعر الحليب لسنتي 1949 - 1950}}$

$$= \frac{413}{\frac{1}{2}\ (395\ +\ 389)} = \frac{413}{392} = 105.4\%$$

منسوب أسعار الزبد $= \dfrac{\text{سعر الزبد في سنة 1958}}{\text{سعر الزبد لسنتي 1949 - 1950}}$

$$= \frac{597}{\frac{1}{2}\ (615\ +\ 622)} = \frac{597}{618.5} = 96.5\%$$

منسوب أسعار الجبن = $\frac{\text{سعر الجبن في سنة 1958}}{\text{سعر الجبن لسنتي 1949 — 1950}}$

$$= \frac{389}{1/2\ (348+354)} = \frac{389}{351} = 110.8\ \%$$

الوسط الحسابي البسيط لمناسيب أسعار منتجات الحليب

$$\frac{105.4 + 96.5 + 110.8}{3} = 104.2\ \%$$

سؤال الوسط المرجح لمناسيب الأسعار

استعمل طريقة الوسط الحسابي المرجح لمناسيب أسعار الانتاج للفحم والبنزين(لدولة ما)في سنتي 1949 و 1958 كما موضحه في الجدول أدناه مستعملا أوزان سنة 1958

الاسعار (بالدولار)			الكميات المنتجة		
	1949	1958		1949	1958
الفحم	20.13	28.20	الفحم	3,559 مليون طن	1,821 مليون طن
البنزين	20.30	21.40	البنزين	80.3 مليون برميل	118,6 مليون برميل

الحل

الوسط الحسابي المرجح لمناسيب الاسعار

$$\frac{\Sigma(p_n/p_0)\ (p_n q_n)}{\Sigma\ p_n q_n} = \frac{\text{مجموع (مناسيب الاسعار) (القيم لسنة 1958)}}{\text{مجموع (القيم لسنة 1958)}}$$

$$= \frac{\frac{28.2}{20.13} \times 28.2 \times 1821000 + \frac{21.4}{20.3} \times 21.4 \times 118600,000}{28.2 \times 1821000 + 21.4 \times 118600\,000}$$

$$= \frac{1,4 \times 51.35 + 1,05 \times 2538,04}{51.35 + 2538.04}$$

$$\frac{71.89 + 2664,94}{2589.39} = \frac{2736.83}{2589.39} = 105\ \%$$

تمــرين (الأرقام القياسيـة) :

السـؤال

الجدول أدناه يوضح أسعار وكميات المعادن (ألمنيوم ، نحاس ، رصاص ، صفيح وزنك) في إحدى الدول الكبرى للسنين 1949 و 1956 و 1957 .

المعدن	الاسعـــــار سنت للرطـــل			الكميـــات (مليون طن)		
	1957	1956	1949	1957	1956	1949
المنيـوم	27.52	26.01	17.00	3698	3707	1357
نحـاس	29.99	41.88	19,36	2478	2734	2144
رصـــاص	14.46	15.81	15,18	2276	2420	1916
صفيـح	96.17	101.26	99.32	186	202	161
زنــك	11.40	13,49	12.15	1424	2018	1872

المطلوب حساب الآتي :

أ) منسوب الأسعار لسنة 1956 باعتبار سنة 1949 سنة الأســاس .

ب) منسوب الأسعار لسنة 1957 باعتبار سنتي 1949 و 1956 كفترة أساس .

جـ) الرقم القياسي التجميعي البسيط لسنة 1957 متخذاً سنة 1956 كسنة الأساس .

د) الوسط البسيط لمناسيب الاسعار .

هـ) رقم (لاسبيرز) القياسي لسنة 1957 متخذاً سنة 1956 كسنة الأساس .

و) رقم (باشي) القياسي لسنة 1957 متخذاً سنة 1956 كسنة الأساس .

ز) رقم (فيشر) القياسي المثالي لسنة 1957 متخذاً سنة 1956 كسنة الأساس .

جـ) رقم (مارشال ـ ادجورث) القياسي لسنة 1957 متخذاً سنة 1956 كسنة الأساس .

ط) رقم (لاسبيرز) القياسي للكميات لسنة 1956 متخذاً سنة 1949 كسنة الأساس .

تمـــاريـــن

الســؤال رقم (1) :

عشر أشخاص يدفع كل منهم30% من دخله الشهري للسكن كما يدفع خمس أشخاص آخرين 25% من دخلهم الشهري للسكن فإذا علمت أن العشر أشخاص المذكورين أولاً يدفعون شهرياً المبالغ التالية للسكن (بالدينار) :

30 ، 33 ، 45 ، 48 ، 51 .

54 ، 36 ، 48 ، 51 ، 57

أمـا الخمس أشخـاص الآخــرين فيدفعــون الإيجــارات الشهــرية الآتية (بالدينار) :

72 ، 80 ، 64 ، 76 ، 84

المطلوب حساب الآتي:

أ ـ الوسط الحسابي للدخل السنوي للخمسة عشر شخصاً.

ب ـ الدخل السنوي الوسيط للخمسة عشر شخصاً.

جـ ـ الدخل السنوي المنوال للخمسة عشر شخصاً.

الســؤال رقم (2) :

الجدول أدناه يوضح الدرجات التي حصل عليها مائة طالـب في الامتحـان بإحدى المدارس الابتدائية.

الدرجة	عدد الطلاب
4.5	2
14.5	3
24.5	6
34.5	13
44.5	22
54.5	24
64.5	16
74.5	8
84.5	5
94.5	1

المطلوب : عمل الآتي:

أ ـ جدول تكراري يبين حدود الفئات .

ب ـ جدول تكراري تجميعي صاعـد .

الســؤال رقم (3) :

بالرجوع الى الجدول التكراري في السؤال الثاني احسب ما يأتي:

أ ـ الوسيــط.

ب ـ الربيع الأول .

جـ ـ الانحراف المعياري .

د ـ التشتت النســبي .

هـ ـ الدرجة المعيارية لطالب حصل على 47 درجة في هذا الامتحان .

السؤال رقم(4) :

لاحظ إداريو احدى شركات الشحن والتفريغ أن تكلفة صيانة جرارات الشحن تزداد مع عمر الجرار . ولدراسة هذه الحالة فقد قامت الشركة بجمع المعلومات الآتية عن عمر الجرار وتكلفة الصيانة .

عمر الجرار (بالسنين)	التكلفة كل 6 شهور (بالدينار)
4.5	60
4.5	100
4.5	110
4.0	50
4.0	70
4.0	60
5.0	90
5.0	150
5.5	100
5.0	120
0.5	20
0.5	20
6.0	80
6.0	140
1.0	100
1.0	50
1.0	60

أ ـ هل تشير هذه البيانات لصحة هذه الملاحظة.

(الإجابة يجب أن تكون باستعمال طريقة إحصائية تحليلية).

ب ـ أوجد خط الانحدار الذين يربط بين هذه المتغيرين.

جـ ـ كم التكلفة (في مدة ستة شهور) لجرار عمره 4 سنوات و6 شهور.

د ـ كم يجب أن يكون عمر الجرار بحيث لا تتعدى كلفة صيانته (كل ستة شهور) 95 دينار.

السؤال رقم (5) :

أ ـ لقد عرفنا تباين البيانات S^2 بالمعادلة التالية : ـ

$$S^2 = \frac{\Sigma\ (X_i - \overline{X})^2}{n - 1}$$

متى تكون قيمة هذا التباين صفراً.

ب ـ أكتب مثالاً لبيانات تتكون من عشرة أرقام بحيث تكون قيمة التباين لهذه البيانات صفراً.

السؤال رقم (6) :

الجدول أدناه يبين أوزان مائة طالب في إحدى المدارس.

الوزن بالكيلو جرام	عدد الطلاب
50 — 52	5
53 — 55	18
56 — 58	42
59 — 61	27
62 — 64	8

المطلوب حساب الآتي:

أ ـ الوسط الحسابي. ب ـ الوسيط. جـ ـ الربيع الثالث.

السؤال رقم (7):

بالرجوع الى الجدول في السؤال السادس ، احسب ما يأتي:

1 ـ التباين للأوزان.

2 ـ التشتت النسبي.

3 ـ ما هي الدرجة المعيارية لطالب وزنه 54 كيلو جراماً.

السؤال رقم (8):

اختير حكمان احمد (X) ومحمد (Y) ليحكما على اللياقة البدنية لعشرة من الرياضيين وكان تقييم هؤلاء العشرة أشخاص كما يلي: ـ

الشخص	الحكم أحمد (X)	الحكم محمد (Y)
1	70	40
2	80	60
3	60	80
4	50	100
5	20	30
6	30	20
7	100	10
8	10	90
9	90	70
10	40	50

لاختبار علاقة الإرتباط بين هذين التقيمين احسب معامل الارتباط (معامل بيرسون).

السـؤال رقم (9):

أ ـ لإيجاد قياس للتشتت حاول أحد الاحصائيين أن يستعمل القانون الآتي للحصول على قيمة للتشتت للبيانات: ـ

$$\text{Dispersion} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})}{n - 1}$$

ولكنه اكتشف بعد لحظة من التفكير بأن هذا القانون لا يصلح لأنه يعطي نفس الرقم للتشتت مهما اختلفت قيم البيانات.

هل هذا صحيح؟ وإذا كان كذلك ما هي هذه القيمة الثابتة للتشتت والتي لا تختلف مهما تغيرت قيم البيانات؟

ب ـ اكتب أي مثال تختاره أنت مكون من ثمانية أرقام بحيث يتساوى فيه الوسط الحسابي والوسيط والمنوال.

السـؤال رقم (10):

الجدول أدناه يبين أطوال مائة مريض في احدى المستشفيات.

الطول (بالبوصة)	عدد الأشخاص
60 — 62	5
63 — 65	18
66 — 68	42
69 — 71	27
72 — 74	8

المطلوب حساب الآتي :

1 ـ الوسط الحسابي.

2 ـ الوسيــط.

3 ـ المنــوال.

السـؤال رقم (11) :

بالرجوع الى الجدول في السؤال رقم (10) ، احسب ما يأتي : ـ

1 ـ الإنحراف المعياري للأطوال.

2 ـ التشتت النسبي للأطــوال.

3 ـ ما هي الدرجة المعيارية لشخص طوله 71 بوصــة.

السـؤال رقم (12) :

أختير حكمان احمد (X) ومحمد (Y) ليحكما على اللياقة البدنية لعشرة من الرياضيين ، وكان ترتيب هؤلاء العشرة أشخاص كما يلي : ـ

الشخص	الحكم احمد (X)	الحكم محمد (Y)
1	7	4
2	8	6
3	6	8
4	5	10
5	2	3
6	3	2
7	10	1
8	1	9
9	9	7
10	4	5

هل هناك ارتباط بين هــذه الرتــب؟

السؤال رقم (13)

الجدول رقم (1) ادناه يوضح اسعار (بالليرة السورية) بيع المنتجات البترولية بالطن في الجمهورية العربية السورية لسنة 1979 و 1980 .

جدول رقم (1)*

المنتج	سعر البيع للمستهلك (ليرة سورية)	
	1979	1980
غاز البترول المسيل	680	680
غازولين ممتاز	1243	1923
غازولين عادي	1096	1803
كيروسين	318	381
زيت الغاز	299	597
زيت وقود ثقيل	281	281
اسفلت	357	357

* (المرجع : التقرير الاحصائي السنوي 1979 —1980 للأوابك) .

اما الجدول رقم (2) فيوضح الكميات المستهلكة لنفس المواد البترولية لنفس السنين في الجمهورية العربية السورية :

الجدول رقم (2)**

المنتج	الاستهلاك المحلي(الف برميل) في اليوم	
	1979	1980
غاز البترول المسيل	3.5	3.8
غازولين ممتاز	13.5	11.3
غازولين عادي	1.2	1
كيروسين	8.3	7
زيت الغاز	50.4	49.9
زيت وقود ثقيل	16.8	17.4
اسفلت	3.7	4.1

ملاحظة : الطن = 6.94 برميلا .

** نفس المصدر .

المطلوب :

حساب ما يأتي : (مع اعتبار سنة 1979 سنة أساس)

1 ـ منسوب اسعار الغازولين الممتاز لسنة 1980 .

2 ـ منسوب اسعار الكيروسين لسنة 1980 .

3 ـ منسوب اسعار الاسفلت لسنة 1979 .

4 ـ الوسط الحسابي البسيط لمناسيب اسعار المواد البترولية .

5 ـ رقم لاسبيرز القياسي للاسعار .

6 ـ رقم باشي القياسي للاسعار .

7 ـ رقم فيشر القياسي للاسعار .

المراجـــع

أولاً: المراجع العربيــة :

(1) د. احمد عباده سرحان، **طرق التحليل الإحصائي**، دار المعارف القاهــرة.

(2) د. عبد العـظيم أنيس، **محـاضرات في الاحصــاء التطبيقـي**، المعهـد العربـي للتخطيط، الكويت 1980

(3) د. فاروق عبد العظيم احمد ، **مقدمة الـطرق الاحصائيـة** ، دار المطبـوعات الجامعية الاسكندرية 1979 .

(4) د. محمد علي بشير، **مقدمة في طرق الاحصاء وتصميم التجارب**، دار المعارف بمصر 1976 .

(5) د. مختار محمود الهانسي، **مقدمة طرق الاحصاء الاجتماعـي**، مؤسســة شبـاب الجامعة ، الاسكندرية.

ثانياً: المراجع الانجليزية

(1) Book, S.A.; **Essentials of Statistics.**
McGraw Hill Book Company, New York, 1978.

(2) Hoel, P. and Jessen, R.; **Basic Statistics for Business and Economics.**
John Wiley and Sons, Inc., New York, 1971.

(3) Mendenhall, W. and Ott, L.; **Understanding Statistics.**
Duxbury Press, Mass. 1976.

(4) Spiegel, M.; **Theories and Problems of Statistics.**
McGraw Hill Press, New York, 1961.

(5) Wonnacott, R.J. and Wonnacott, T.H.; **Introductory Statistics.**
John Wiley and Sons, New York, 1976.

تصويب الاخطاء

رقم الصفحة	السطر	الخطـــأ	الصـــواب
٣٤	السطر قبل الاخير	قيمتها مساوية واكبر	قيمتها مساوية أو اكبر
٣٧	المحور الافقي في الشكل	درجات الامتحا	درجات الامتحان
٦٣	السطر الاول	الموضع النسبي للوسيط والوسيط والمنوال	الموضع النسبي للوسط ،والوسيط والمنوال
	في الشكل الاول من اليمين	$M+26 \quad M \quad M+36$	$M-36 \quad M \quad M+36$
٧٨	السطر ٩	بين نفس الرقمين أى بين 200 و 280	بين نفس الرقمين $(X+26 \,،\, X+26)$ أى بين 200 و 280
٨٥	السطر ١٥	بالتعويض لقيم المتوسط الحسابي	بالتعويض لقيم الوسط الحسابي
٨٦	اسفل جدول	$118-33\bar{X}$	$118-33=\bar{X}$
١٠١	السطر ٣	$a=\hat{\alpha}=\bar{y}-b\bar{X}$	$a=\hat{\alpha}=\bar{Y}-b\bar{X}$
	السطر ٤	$b=\hat{\beta}=\frac{Sxy}{S^2x}$	$b=\hat{\beta}=\frac{Sxy}{S^2x}$
	السطر قبل الاخير	هناك ارتباطا خطيا بين X , y	هناك ارتباطا خطيا تاما بين X , y
١٠٢	السطر ١١	وبشكل عام فان Sx-y يساوى Sy-x	وبشكل عام فان Sx-y لا يساوى Sy-x
١٠٣	السطر ٢	(16) $\Sigma(Y_i-\hat{Y})^2=\Sigma(Y-\hat{Y})^2+\Sigma(\hat{Y}-\bar{Y})^2$	(16) $\Sigma(Y_i-\hat{Y}_i)^2=\Sigma(Y-\hat{Y})^2+\Sigma(\hat{Y}-\bar{Y})^2$
	السطر ٥	$yi-yi$ تسلك سلوكا عشوائيا	$yi-\hat{y}i$ تسلك سلوكا عشوائيا
١٠٤	السطر ٣	في حالة الارتباط الخطي فان تظل كما هي	في حالة الارتباط الخطي فان r تظل كما هي
١١٦	السطر ١٥	بما ان	د ــ بما ان
١٢٢	السطر ١٠	الصيغة (٢٧) تسمي	الصيغة (32) تسمي
١٣٣	السطر ١٠	السنوات 1955, 1960	السنوات 1960, 1955
١٣٧	السطر ١٠	طريقة سنة الاساس	طريقة سنه الاساس
١٤٠	السطر ١٥	في المعادلات من (8) الى (13) نحصل	في المعادلات من (8) الى (13) نحصل
	السطر الاخير	$\frac{\Sigma q_0 p_n}{\Sigma q_0 P_0}$	$\frac{\Sigma p_n q_n}{\Sigma p_0 q_n}$
١٤٩	السطر ٨	$104-4\% = \frac{4411226,81}{4222063,42}$	$104-4\% = \frac{4411226-81}{4222063-42}$
	السطر ١٣	$\sqrt{104\ 5\times104.4} =$	$\sqrt{104{\cdot}5\times104.4} =$
١٥٠	السطر الاول	$M\text{-}E=\frac{\Sigma p_n(q_0+q_n)}{\Sigma p_0(q_0+q_n)}$ د ـ رقم مارشال ارجورث القياس	رقم مارشال ـ ارجورث القياس $M\text{-}E=\frac{\Sigma p_n(q_0+q_n)}{\Sigma p_0(q_0+q_n)}$


Bibliotheca Alexandrina
0332626